

The Psychologist as Detective

An Introduction to Conducting
Research in Psychology

Fifth Edition

Randolph A. Smith

Lamar University

Stephen F. Davis

Morningside College

Prentice Hall

Upper Saddle River London Singapore
Toronto Tokyo Sydney Hong Kong Mexico City

We dedicate the fifth edition of this text to the loving memory of Corliss Smith (1952–2009).
Her support made this book possible. Please consider a donation to the ALS Association
(www.alsa.org) to fight this cruel disease.

Editorial Director: Leah Jewell
Editor in Chief: Jessica Mosher
Executive Editor: Jeff Marshall
Editorial Project Manager: LeeAnn Doherty
Editorial Assistant: Amy Trudell
Director of Marketing: Brandy Dawson
Marketing Manager: Nicole Kunzmann
Assistant Marketing Manager: Laura Kennedy
Marketing Assistant: Jennifer Lang
Managing Editor: Maureen Richardson
Project Liaison: Shelly Kupperman
Senior Operations Supervisor: Sherry Lewis
Director, Cover Design: Jayne Conte
Cover Designer: Lisbeth Axell

AV Project Manager: Maria Piper
Manager, Visual Research: Beth Brenzel
Manager, Rights and Permissions: Zina Arabia
Image Permissions Coordinator:
Angelique Sharps
Manager, Cover Visual Research and Permissions:
Karen Sanatar
Cover Art: Boris Lyubner/Illustration Works/Corbis
Media Director: Paul DeLuca
Supplements Editor: Ginny Livsey
Composition/Full-Service Project Management:
Macmillan Publishing Solutions/Laura Lawrie
Printer/Binder: Edwards Brothers
Cover Printer: Coral Graphics

This book was set in 9/12 Poppl-Laudatio Light.

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on appropriate page within text.

Copyright © 2010, 2007, 2004, 2001, 1997 Pearson Education, Inc., publishing as Prentice Hall, 1 Lake Street, Upper Saddle River, NJ 07458. All rights reserved. Manufactured in the United States of America. This publication is protected by Copyright, and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. To obtain permission(s) to use material from this work, please submit a written request to: Pearson Education, Inc., Permissions Department, 1 Lake Street, Upper Saddle River, NJ 07458.

Library of Congress Cataloging-in-Publication Data

Smith, Randolph A.

The psychologist as detective : an introduction to conducting research in psychology / Randolph A. Smith, Stephen F. Davis. — 5th ed.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-205-68740-4 (alk. paper)

1. Psychology—Research—Textbooks. 2. Psychology—Research—Methodology—Textbooks. 3. Psychology, Experimental—Textbooks. I. Davis, Stephen F. II. Title.

BF76.5.S54 2010

150.72—dc22

2009001784

10 9 8 7 6 5 4 3 2 1

Prentice Hall
is an imprint of



www.pearsonhighered.com

ISBN 10: 0-205-68740-7
ISBN 13: 978-0-205-68740-4

Contents

Preface xii

CHAPTER 1 Psychological Research and the Research Methods Course 1

How Psychologists Acquire Knowledge 3

The Research Process 4

Finding a Problem 5

Reviewing the Literature 5

Theoretical Considerations 5

Hypothesis 6

Research Plan 6

Conducting the Research Project 6

Analysis of Research Findings 6

Decisions in Terms of Past Research and Theory 7

Preparing the Research Report 7

Sharing Your Results: Presentation and Publication 7

Finding a New Problem 11

Why Is the Research Methods Course Important? 11

Review Summary 12

Check Your Progress 13

Key Terms 13

Looking Ahead 13

CHAPTER 2 Developing a Good Research Idea and Conducting an Ethical Project 14

The Research Idea 14

Characteristics of Good Research Ideas 14

Sources of Research Ideas 16

Developing a Research Question 19

Surveying the Psychological Literature 20

Review Summary 28

Check Your Progress 28

The Need for Ethical Principles 29

APA Principles in the Conduct of Research with Humans	32
Is Deception in Research Necessary?	34
Informed Consent	34
Participants at Risk and Participants at Minimal Risk	36
Vulnerable Populations	36
The Debriefing Session	37
The Ethical Use of Animals in Psychological Research	38
The Institutional Review Board	39
The Experimenter's Responsibility	40
The Participants' Responsibility	40
The Researcher's Ethical Obligations Once the Research Is Completed	42
Plagiarism	42
Fabrication of Data	43
Lying with Statistics	45
Citing Your References Correctly	45
Review Summary	47
Check Your Progress	47
Key Terms	48
Looking Ahead	48

CHAPTER 3 Qualitative Research Methods 49

General Overview	50
Characteristics of Qualitative Research	50
Data Analysis	51
Selected Examples of Qualitative Research Methods	52
Naturalistic Observation	52
Ethnographic Inquiry	52
Focus Groups	53
Case Studies	54
Grounded Theory	54
Participatory Action Research	56
Review Summary	57
Check Your Progress	57
Key Terms	58
Looking Ahead	58

CHAPTER 4 Nonexperimental Methods: Descriptive Methods, Correlational Studies, Ex Post Facto Studies, Surveys and Questionnaires, Sampling, and Basic Research Strategies 59

Descriptive Methods	59
Archival and Previously Recorded Sources of Data	59

Naturalistic Observation 61
Choosing Behaviors and Recording Techniques 63

Correlational Studies 66
The Nature of Correlations 66
Correlational Research 67
Review Summary 68
Check Your Progress 69

Ex Post Facto Studies 69

Surveys, Questionnaires, Tests, and Inventories 70
Surveys and Questionnaires 70
Tests and Inventories 76

Sampling Considerations and Basic Research Strategies 79
Sampling 79
Basic Research Strategies 81
Review Summary 83
Check Your Progress 84
Key Terms 85
Looking Ahead 85

CHAPTER 5 Using the Scientific Method in Psychology 86

Components of the Scientific Method 86
Objectivity 86
Confirmation of Findings 87
Self-Correction 87
Control 87

The Psychological Experiment 88
Independent Variable 88
Dependent Variable 89
Extraneous Variables 89

Establishing Cause-and-Effect Relations 89

Formulating the Research Hypothesis 91

Characteristics of the Research Hypothesis 92
Types of Statements 92
Types of Reasoning 94
A New View of Hypothesis Testing 95
Directional Versus Nondirectional Research Hypotheses 96
Review Summary 97
Check Your Progress 98
Key Terms 99

CHAPTER 6 Conducting a Good Experiment I: Variables and Control 100

The Nature of Variables	100
Operationally Defining Variables	101
Independent Variables	102
Types of IVs	102
Extraneous Variables (Confounders)	103
Dependent Variables	105
Selecting the DV	106
Recording or Measuring the DV	106
Recording More Than One DV	107
Characteristics of a Good DV	108
Nuisance Variables	109
Review Summary	111
Check Your Progress	112
Controlling Extraneous Variables	112
Basic Control Techniques	113
Review Summary	124
Check Your Progress	124
Key Terms	125
Looking Ahead	125

CHAPTER 7 Conducting a Good Experiment II: Final Considerations, Unanticipated Influences, and Cross-Cultural Issues 126

Participants	126
Types of Participants	126
Number of Participants	128
Apparatus	129
IV Presentation	129
DV Recording	131
Review Summary	131
Check Your Progress	132
The Experimenter as an Extraneous Variable	132
Experimenter Characteristics	133
Experimenter Expectancies	133
Controlling Experimenter Effects	134
Participant Perceptions as Extraneous Variables	135
Demand Characteristics and Good Participants	135
Response Bias	137
Controlling Participant Effects	138

Review Summary 140
Check Your Progress 140

The Interface Between Research and Culture 141

Culture, Knowledge, and Truth 142
The Effect of Culture on Research 143
Methodology and Analysis Issues 143
Review Summary 145
Check Your Progress 146
Key Terms 146
Looking Ahead 146

CHAPTER 8 Internal and External Validity 147

Internal Validity: Evaluating Your Experiment from the Inside 147

Threats to Internal Validity 147
Protecting Internal Validity 155
Review Summary 155
Check Your Progress 156

External Validity: Generalizing Your Experiment to the Outside 157

Threats to External Validity (Based on Methods) 159
Threats to External Validity (Based on Our Participants) 162
The Devil's Advocate: Is External Validity Always Necessary? 167
Review Summary 169
Check Your Progress 170
Key Terms 170
Looking Ahead 170

CHAPTER 9 Using Statistics to Answer Questions 171

Descriptive Statistics 171
Scales of Measurement 172
Measures of Central Tendency 173
Graphing Your Results 176
Calculating and Computing Statistics 181
Measures of Variability 182
Review Summary 187
Check Your Progress 187

Correlation 188

The Pearson Product-Moment Correlation Coefficient 191

Inferential Statistics 192

What Is Significant? 192
The *t* Test 193
One-Tail Versus Two-Tail Tests of Significance 196
The Logic of Significance Testing 196
When Statistics Go Astray: Type I and Type II Errors 199

Effect Size 200

- Review Summary 201
- Check Your Progress 201
- Key Terms 202
- Looking Ahead 202

CHAPTER 10**Designing, Conducting, Analyzing, and Interpreting Experiments with Two Groups 203****Experimental Design: The Basic Building Blocks 203**

- The Two-Group Design 204
- Review Summary 213
- Check Your Progress 214
- Comparing Two-Group Designs 214
- Variations on the Two-Group Design 218
- Review Summary 220
- Check Your Progress 220

Statistical Analysis: What Do Your Data Show? 220

- The Relation Between Experimental Design and Statistics 220
- Analyzing Two-Group Designs 221
- Calculating Your Statistics 221

Interpretation: Making Sense of Your Statistics 222

- Interpreting Computer Statistical Output 222

The Continuing Research Problem 228

- Review Summary 229
- Check Your Progress 230
- Key Terms 230
- Looking Ahead 230

CHAPTER 11**Designing, Conducting, Analyzing, and Interpreting Experiments with More Than Two Groups 231****Experimental Design: Adding to the Basic Building Block 231**

- The Multiple-Group Design 232
- Comparing the Multiple-Group and Two-Group Designs 238
- Comparing Multiple-Group Designs 240
- Variations on the Multiple-Group Design 241
- Review Summary 242
- Check Your Progress 243

Statistical Analysis: What Do Your Data Show? 243

- Analyzing Multiple-Group Designs 243
- Planning Your Experiment 244
- Rationale of ANOVA 245

Interpretation: Making Sense of Your Statistics 247**Interpreting Computer Statistical Output 248****The Continuing Research Problem 256****Review Summary 257****Check Your Progress 257****Key Terms 258****Looking Ahead 258****CHAPTER 12****Designing, Conducting, Analyzing, and Interpreting Experiments with Multiple Independent Variables 259****Experimental Design: Doubling the Basic Building Block 259****The Factorial Design 260****Review Summary 272****Check Your Progress 273****Comparing the Factorial Design to Two-Group and Multiple-Group Designs 273****Choosing a Factorial Design 276****Variations on Factorial Designs 277****Review Summary 282****Check Your Progress 282****Statistical Analysis: What Do Your Data Show? 283****Naming Factorial Designs 283****Planning the Statistical Analysis 284****Rationale of Factorial ANOVA 284****Understanding Interactions 285****Interpretation: Making Sense of Your Statistics 287****Interpreting Computer Statistical Output 287****A Final Note 296****The Continuing Research Problem 296****Review Summary 297****Check Your Progress 298****Key Terms 299****Looking Ahead 299****CHAPTER 13****Alternative Research Designs 300****Protecting Internal Validity Revisited 300****Examining Your Experiment from the Inside 300****Protecting Internal Validity with Research Designs 303****Conclusion 307****Review Summary 307****Check Your Progress 308**

Single-Case Experimental Designs	308
History of Single-Case Experimental Designs	310
Uses of Single-Case Experimental Designs	310
General Procedures of Single-Case Experimental Designs	311
Statistics and Single-Case Experimental Designs	313
Representative Single-Case Experimental Designs	314
Review Summary	320
Check Your Progress	321
Quasi-Experimental Designs	321
History of Quasi-Experimental Designs	322
Uses of Quasi-Experimental Designs	323
Representative Quasi-Experimental Designs	323
Review Summary	330
Check Your Progress	331
Key Terms	332
Looking Ahead	332

CHAPTER 14 Writing and Assembling an APA-Format Research Report 333

What Is APA Format?	333
Sections of the APA-Format Paper	334
Title Page	335
Abstract	336
Introduction	338
Method	343
Results	348
Discussion	353
References	356
Appendix	362
Author Note	362
Headings	363
Review Summary	370
Check Your Progress	370
Writing in APA Style	371
General Guidelines	372
Grammatical Guidelines	373
APA Editorial Style	376
Preparing Your Manuscript	376
Student Views of Professional Activities	379
Review Summary	385
Check Your Progress	386
Key Terms	387
Looking Ahead	387

- Appendix A: Statistical Tables 389**
- Appendix B: Selected Statistical Formulae 395**
- Appendix C: Factorial Design with Three Independent Variables 397**
- Appendix D: Check Your Progress Answers 400**
- Glossary 417**
- References 427**
- Index 435**

Preface

Because our objectives and goals and the specific features we've included in *The Psychologist as Detective* remain unchanged in this fifth edition, our original Preface follows.

Note to the Instructor

Margery Franklin (1990) quoted former Clark University professor and chair Heinz Werner's views on psychological research. Werner indicated:

I got rather apprehensive at finding that students were frequently taught that there was only one acceptable way of conduct in the laboratory—there has to be an hypothesis set up, or a set of hypotheses, and the main job of the experimenter is to prove or disprove the hypothesis. What is missed here is the function of the scientist as a discoverer and explorer of unknown lands. . . . Hypotheses . . . are essential elements of inquiry, but they are so not as rigid propositions but as flexible parts of the process of searching; by the same token, conclusions drawn from the results are as much an end as a beginning. . . . Now . . . academic psychologists [are beginning] to see research not as a rigid exercise of rules of a game but as a problem solving procedure, a probing into unknown lands with plans which are not fixed but modifiable, with progress and retreat, with branching out into various directions or concentration on one. (p. 185)

Clearly Werner's views are as applicable in the 21st century as they were during the heyday of behaviorism; they reflect perfectly the intent of this text.

From our vantage point, research in psychology is like a detective case; hence the title we have chosen, *The Psychologist as Detective*. A problem presents itself; we discover clues; we must evaluate bits of evidence that compete for our attention and accept or discard them; and finally, we prepare a report or summary of the case (research) for consideration by our peers.

When presented in this light, the research process in psychology will, we believe, be an interesting and stimulating endeavor for students. In short, our goal is to attract students to psychological research because of its inherent interest.

To accomplish this goal, we have incorporated several pedagogical features in this text:

1. To provide a sense of relevance and continuity, the theme of "psychologist as detective" runs throughout the text.
2. **Interactive Style of Writing.** Because we believe that the experimental psychology/research methods text should be lively and engaging, we employ an interactive, conversational style of writing that we hope will help draw students into the material.
3. **The Psychological Detective Feature.** The questions or situations posed by these sections that appear throughout each chapter will encourage students to engage in critical thinking exercises. These sections also serve as excellent stimulants for productive class discussions.

4. **Marginal Definitions.** Key definitions appear in the margin, close to the introduction of the term in the text.
5. **Review Summaries.** To help students master smaller chunks of material, each chapter contains one or more review summaries.
6. **Check Your Progress.** A Check Your Progress feature follows each Review Summary. Students can use these sections to test their mastery of the material they have just completed. These study breaks should be especially helpful to your students when they prepare for quizzes and examinations.

We hope that these special features will provide your students with a positive experience as they learn the fundamentals of research methodology in psychology.

Note to the Student

Welcome to the world of psychological research! Because the two of us have taught this course for over 60 years (combined!), we have seen the excitement that research can generate in student after student. As you will learn, conducting psychological research is very much like being a detective on a case.

Throughout this text we have tried to make it clear that research is something that you can (and should) become involved in. We hope you will enjoy reading about the student projects that we use as research examples throughout this text. Student research projects are making valuable contributions to our field. We hope to see your name among those making such contributions!

At this point we encourage you to stop immediately to review the list of pedagogical features highlighted in the “Note to the Instructor.”

Did you humor us by actually looking at that list? If not, please do so now. To make full use of this text, you need to become actively involved; these pedagogical features will help you. Active involvement means that you need to stop and think about The Psychological Detective sections immediately when you encounter them, refer to figures and tables when directed to do so, and complete the Check Your Progress sections when they appear. Becoming actively involved in this course helps the material come alive; your grade and your future involvement in psychology will thank you.

What's New for the Fifth Edition?

We are excited about the continuing evolution of *The Psychologist as Detective: An Introduction to Conducting Research in Psychology*. We have implemented several major changes and additions for the fifth edition:

1. Based on feedback from adopters, we have:
 - A. Streamlined the first chapter and sharpened the focus to deal with the general topic of research methods.
 - B. Devoted a complete chapter (Chapter 3) to the coverage of Qualitative Research Methods and combined our coverage of other nonexperimental research methods into one, separate chapter.
 - C. Created a separate chapter to introduce students to the basics of using the scientific methods in psychological research (Chapter 5).

2. An updated and expanded Instructor's Manual and Electronic Test Item File are also available to accompany *The Psychologist as Detective*. Please contact your local Prentice Hall representative or log onto www.pearsonhighered.com.

Acknowledgments

We express our appreciation to the consultants who suggested improvements for this text; their comments were especially helpful: Bryan K. Saville, James Madison University; Constance Jones, California State University, Fresno; Michael T. Dreznick, Our Lady of the Lake College; Jonathan W. Amburgey, University of Utah; Nancy J. Knous, Northwestern Oklahoma State University; Lauren Mizock, Suffolk University; Roger A. Chadwick, New Mexico State University; Katia Shkurkin, Saint Martin's University; Oriel Strickland, California State University, Sacramento; Keiko Taga, California State University, Fullerton.

And from previous edition reviews:

Chris Spatz, Hendrix College; Beth Dietz Uhler, Miami University (Ohio); Janet Larsen, John Carroll University; Doreen Arcus, University of Massachusetts, Lowell; Lynette Zeleny, California State University, Fresno; Robert Batsell, Kalamazoo College, Scott Gronlund, University of Oklahoma; Laura Bowman, Central Connecticut State University; Celia Brownell, University of Pittsburgh; Lee Fernandez, Yosemite Community College; Michael Brannick, University of Southern Florida; Maureen McCarthy, Kennesaw State University; Elizabeth Yost Hammer, Loyola University; Terry Pettijohn, Mercyhurst College; Gina Grimshaw, California State University, San Marcos. Susan Burns, Morningside College; Mark Stellmack, University of Minnesota; Karen Schmidt, University of Virginia; Morton Heller, Eastern Illinois University; Joy Drinon, Milligan College; Rebecca Regeth, California University of Pennsylvania; Christina Sinisi, Charleston Southern University; Rick Froman, John Brown University; Libbe Gray, Austin College; Mizuho Arai, University of Massachusetts, Boston; Theresa L. White, Le Moyne College; David E. Anderson, Allegheny College; Jerome Lee, Albright College; Scott A. Bailey, Texas Lutheran University.

R. A. S.

S. F. D.

Psychological Research and the Research Methods Course

CHAPTER

1

How Psychologists Acquire Knowledge

The Research Process

- Finding a Problem • Reviewing the Literature
- Theoretical Considerations • Hypothesis
- Research Plan • Conducting the Research Project
- Analysis of Research Findings

- Decisions in Terms of Past Research and Theory
- Preparing the Research Report
- Sharing Your Results: Presentation and Publication
- Finding a New Problem

Why Is the Research Methods Course Important?

Welcome to the world of psychological research! Because the two of us have taught this course for over 60 years (combined!), we have seen the excitement that research can generate in student after student. Throughout this text we have tried to make it perfectly clear that research is something that you can (and should) become involved in. We hope you will enjoy reading about the student projects we use as research examples throughout this text. Student research projects are making valuable contributions to our field, and we hope to see your name among those making such contributions!

To make full use of this text, you need to become actively involved. Active involvement means that you need to

1. Stop and think about **The Psychological Detective** sections as soon as you encounter them. Each of these sections asks you to think about a question concerning psychological research. Take full advantage of these sections; we designed them to help you think critically about psychological research. Critical thinking is vital to good detective work; we want you to become the best psychological detective possible.
2. Refer to figures and tables when directed to do so.
3. Complete the **Check Your Progress** features when they appear.

Becoming actively involved in this course helps the material come alive; your grade and your future involvement in psychology will reflect your efforts.

We purposely titled our text *The Psychologist as Detective* to convey the excitement and thrill experienced by researchers when they investigate questions that are at the core of what it means to be a psychologist. To make the fifth edition of *The Psychologist as Detective* as lively as possible, we've included several quotations from the world's most famous detective, Sherlock Holmes. For example, Holmes reflected his passion for his work when he said, "I swear that another day shall not have passed before I have done all that man can do to reach the heart of the mystery" (Doyle, 1927, p. 732). Psychological researchers are just as passionate in their pursuit of the truth.

The parallels between conducting psychological research and working on a detective case are striking. The detective has to know the boundaries of the case (the experimenter develops a research question). The detective eliminates suspects (the researcher exercises control over unwanted factors), gathers evidence (the researcher conducts the experiment and makes observations), proposes a solution (the researcher analyzes research results and offers an interpretation), and offers a solution to the jury (researchers share their results and interpretations with their peers).

Our examination of psychological research begins by considering a research project that would intrigue even the best psychological detective:

You have been receiving junk e-mails inviting you to try out the many online dating services for several months. Your typical response is to click the “delete” key; however, you finally start to wonder about online dating services. What characteristics do people look for in a potential date? Are there differences in the characteristics looked for by men and women?

Such questions prompted Kim Driggers and Tasha Helms, students at Oklahoma State University, to conduct a research project that involved an evaluation of salary as a factor in dating. Driggers and Helms (2000) reported, “Male and female college students viewed pictures of the opposite sex and rated the target’s attractiveness and their own willingness to date the target” (p. 76). As you can see from Figure 1-1, they found that willingness to date increased as salary increased. Moreover, the female participants seemed to be more influenced than the male participants by the highest salary level.

For the time being we will not concern ourselves with exactly how Driggers and Helms gathered their information concerning willingness to date. Our concern at present is why they gathered this information. The answer really is quite straightforward—they had an interesting

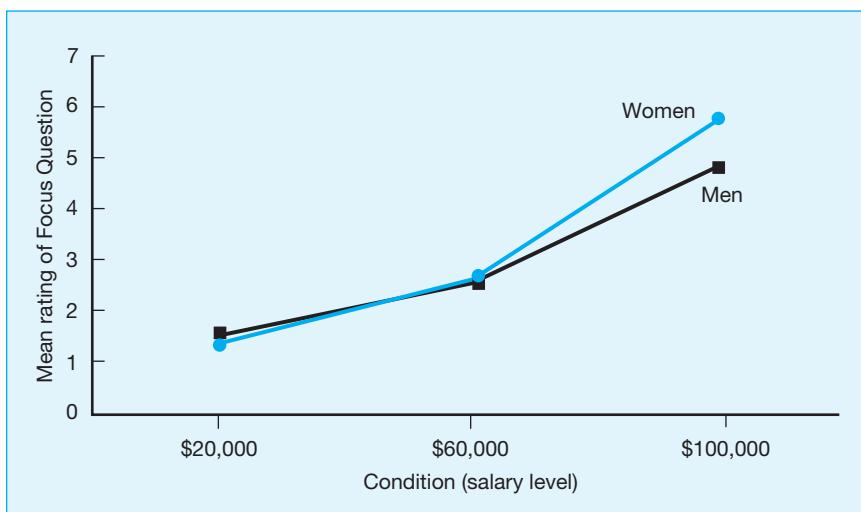


FIGURE 1-1 The Mean Rating of Male and Female Participants for Each Condition (\$20,000, \$60,000, \$100,000) on the Focus Question (“How willing would you be to go on a date with this person?”).

Source: Fig. 1 from Driggers, K. J. and Helms, T. (2000). “The effects of salary on willingness to date.” *Psi Chi Journal of Undergraduate Research*, 5, 76–80. G144. Copyright © 2000 Psi Chi, The National Honor Society in Psychology (www.psichi.org). Reprinted by permission. All rights reserved.

question that needed an answer. Asking and then attempting to answer questions are at the heart of what psychologists do. In short, psychologists are in the business of acquiring new knowledge.

How Psychologists Acquire Knowledge

Part of the reason that psychology always is exciting and interesting is because it deals with the full range of the behavior of humans and animals. Without question, we are talking about an immense number of behaviors. Now, add the vast number of situations and circumstances under which these behaviors can occur and the total number of questions about behavior that psychologists can be interested in is truly staggering.

Because of this great diversity of questions, psychologists have developed different types of research methods. Each method is designed to provide an answer to a research question about a specific behavior in a specific setting. For example, the method researchers use to investigate the effects of a person's salary on desirability as a date likely will be different from the method used to study road rage or the behavior of a white rat in a maze.

In the following chapters we will explore several of the most popular research methods that psychologists have developed. More specifically, the procedures that we will examine include qualitative research; descriptive methods; correlational studies; surveys, questionnaires, tests, and inventories; *ex post facto* studies; and experiments. Table 1-1 provides a brief description of each of these methods and where you will find them in this book.

TABLE 1-1 Research Methods Used by Psychologists

Method	Description	Location
Qualitative Research	A holistic approach that is conducted in a natural setting. It focuses on an entire phenomenon, subculture, or culture. The goal of this approach is to develop a complete description of the behavior of interest.	Chapter 4
Descriptive Methods	Research methods that do not involve the manipulation of factors or variables by the researcher. The researcher may collect data from archives and other previously recorded sources, case studies, and clinical observation.	Chapter 5
Correlational Studies	A mathematical approach that studies the strength of the relation between two factors or variables of interest.	Chapter 5
Surveys, Questionnaires, Tests, and Inventories	An approach that researchers use to assess attitudes, thoughts, emotions, and feelings.	Chapter 6
Ex Post Facto Studies	As the name indicates, the factors or variables of research interest are studies "after the fact" (<i>ex post facto</i>)	Chapter 6
Experiments	An attempt to determine the cause-and-effect relations that exist in nature. This approach involves the manipulation of an independent variable (IV), recording of changes in a dependent variable (DV), and control of extraneous variables.	Chapter 7

As you will see in the following chapters, each of these general methods can have several types or variations. However, regardless of the method that the researcher selects, the general research process will be the same.

The Research Process

A number of interrelated activities make up the research process. These activities appear in Table 1-2. As you can see, one activity leads to the next one until we share our research information with others and the process starts all over again. We will now briefly describe each of these steps but will cover them in greater detail in the following chapters.

TABLE 1-2 Components of the Research Process

Each component builds on the previous components until a new problem is developed and the process starts over again.

Problem	You detect a gap in the knowledge base or wonder about a relation.
Literature Review	Consulting previous reports determines what has been found in the research area of interest.
Theoretical Considerations	The literature review highlights theories that point to relevant research projects.
Hypothesis	The literature review also highlights hypotheses (statements of the relation between variables in more restricted domains of the research area). Such hypotheses will assist in the development of the experimental hypothesis—the predicted outcome of your research project.
Research Plan	You develop the general plan or research design for conducting the research project.
Conducting the Project	You conduct the research project according to the research plan or experimental design.
Analysis of Research Findings	You analyze your research findings. Many projects will involve statistical analyses and statistical decisions.
Decisions in Terms of Past Research and Theory	Your findings guide decisions concerning the relation of the present research project to past research and theoretical considerations.
Preparation of the Research Report	You write a research report describing the rationale, conduct, and results of the project according to accepted American Psychological Association (APA) format.
Sharing Your Results: Presentation and Publication	You share your research report with colleagues at a professional society meeting and/or by publication in a professional journal.
Finding a New Problem	Your research results highlight another gap in our knowledge base, and the research process begins again.

Finding a Problem

Each research project begins as a problem or a question for which we are seeking an answer. For example, Driggers and Helms (2000) wanted to know whether a person's salary affected the willingness of others to date that person.

Reviewing the Literature

Once you have chosen the problem you plan to research, you must discover what psychologists already know about the problem. Thus, the next step is to find out what research studies already exist in this area.



Why is it important to conduct a thorough review of the literature before conducting your research project? Give this question some thought and write down your reasons before reading further.

You may find that the exact project you have in mind has been conducted many times before. Hence, a modification of your idea, not a replication, may be more informative.

Theoretical Considerations

In the course of your literature review you will undoubtedly come across theories that researchers have developed in the area you have chosen to research. A **theory** is a formal statement of the relation among the variables or factors of interest in a particular research area. A theory is not based on the researcher's guess about the relations among variables. The results of numerous research studies typically form the basis for a psychological theory.

Theory A formal statement of the relations among the variables in a given area of research.

Leon Festinger's (1957) cognitive dissonance theory is a good example of a psychological theory that has generated considerable research. Festinger proposed that tension is aroused when two beliefs, thoughts, or behaviors are psychologically inconsistent (dissonant). In turn, we are motivated to reduce this cognitive dissonance by altering our thoughts or behaviors to make them more compatible. For example, (1) believing that high cholesterol is bad for your health and (2) eating pizza (which raises cholesterol) almost every day are inconsistent. The dissonance created by having this incompatible belief and behavior might be reduced by deciding that the reports on the harmful effects of cholesterol really are not correct or by eating pizza less often. Many researchers have tested predictions from Festinger's theory over the years.

All good theories share two common properties. First, they represent an attempt to organize a given body of scientific data. If a theory has not been developed in a particular area of research, we are faced with the task of having to consider the results of many separate experiments and decide how these results are related to each other.

The second property shared by theories is their ability to point the way to new research. By illuminating the relations among relevant variables, a good (i.e., testable) theory also suggests what might logically happen if researchers manipulate these variables in certain

ways. You can think of a theory as being like a road map of your home state. The roads organize and show the relations among the towns and cities. By using this map and a bit of logic, you should be able to get from point A to point B. Thus, the theories that you encounter while conducting a literature review will help point you to a relevant research project.

Hypothesis

If you think a theory is like a road map of your home state, then you can think of a hypothesis as being like the map of a specific town in your state. The

hypothesis attempts to state specific relations among variables within a selected portion of a larger, more comprehensive research area or theory. Within the general domain of cognitive dissonance theory, a number of studies have been concerned with just the finding that cognitive dissonance results in arousal. The hypothesized arousal occurs through increases in such physiological reactions as perspiration and heart rate (Losch & Cacioppo, 1990). Just as the map of your hometown may show you several ways to arrive at your destination, researchers may find that there is more than one route to their research objective. Hence,

researchers may develop several hypotheses to answer the research question. For example, you might predict that a reduction in arousal would result in a decrease in cognitive dissonance. This prediction has been tested; one experiment showed that participants who consumed alcohol (arousal reduction) had reduced levels of cognitive dissonance (Steele, Southwick, & Critchlow, 1981).

As your research project takes shape, you will develop a specific hypothesis. This hypothesis, frequently called the **research or experimental hypothesis**, will be the predicted outcome of your research project. In stating this hypothesis you are stating a testable prediction about the relations between the variables in your project. Based on the scientific literature you have reviewed, your experimental hypothesis will be influenced by other hypotheses and theories that researchers have proposed in your area of interest. For example, your hypothesis

might be "If potential customers dress in old, worn-out clothes, then they will not be waited on as quickly as customers dressed in better clothing."

Research Plan

Once you have formulated your hypothesis, you need a general plan or experimental design for conducting your research and gathering data. This plan is called a **research design**.

Conducting the Research Project

The next step is to conduct the research project. It is not a foregone conclusion that you will conduct your research in a laboratory. You may find yourself gathering research data in a shopping mall, an animal observation station, an archive, or hundreds of other possible locations. You will put all your preparations and controls to the test as you gather your research data.

Analysis of Research Findings

Your research project is not complete when you have gathered the data; the next step is to analyze the data that you gathered. As you will see in our discussion of qualitative research

Hypothesis An attempt to organize certain data and specific relations among variables within a specific portion of a larger, more comprehensive theory.

Research or experimental hypothesis The experimenter's predicted outcome of a research project.

Research design The general plan for conducting research and gathering data.

(see Chapter 4), this analysis may result in a lengthy narrative description of the behavior of interest. However, in many projects numerical data are involved. This is where statistics come into play; statistics is the branch of mathematics that we will use to make sense of and analyze numerical data. Based on the results of our data analysis, we will decide the importance (*significance*) of our research findings.

Decisions in Terms of Past Research and Theory

Once you have conducted your analyses, then you must interpret your results in light of past research and theory. Was your research hypothesis supported? Do your results agree with past research? How do they fit into the current theoretical structure in this research area? If your results do not fit perfectly, what changes need to be made in your interpretation or existing theory to accommodate them? Does lack of support for your hypothesis disconfirm the theory? We will have more to say about hypothesis testing and theories in subsequent chapters.

Researchers want to be able to extend or generalize their results as widely as they legitimately can. Would Driggers and Helms's (2000) results on the effect of salary level and willingness to date also generalize to other types of participants besides college students? This is the type of issue with which generalization deals.

Preparing the Research Report

Before you share the results of your research with the scientific community, you must prepare a written research report. You will prepare this research report according to the format prescribed by the American Psychological Association (APA). This format, often called APA format, is detailed in the *Publication Manual of the American Psychological Association* (2001).

Although many specific details of APA format have evolved over the years, University of Wisconsin psychologist Joseph Jastrow originally proposed the first structure of a research report in the early part of this century (Blumenthal, 1991). Jastrow's purpose in suggesting a standard format for all psychological papers was to make the communication of research results easier and more consistent. A standard form enables researchers to know exactly what to include in their papers and readers to know where to look for specific experimental details, procedures, and results. We will discuss APA format in detail in Chapter 14. We encourage you to look at that chapter now and refer to it throughout the course; the more familiar you are with this format, the easier it will be for you to prepare your own report.

Sharing Your Results: Presentation and Publication

Once you have conducted the project, analyzed the findings, and prepared the research report, it is time to share your results. The two most popular ways to accomplish this objective are (a) to present an oral paper or a poster at a psychological convention and (b) to publish an article in a professional journal.

Even though many of you may be shaking your heads and saying, "I could never do that in a million years," we believe (and know from experience) such accomplishments are within the grasp of most motivated undergraduate psychology students. In fact, such opportunities, especially for presenting papers and posters at psychological conventions, have increased dramatically in recent years. Paramount among these opportunities are a growing number of state and regional student psychology conventions. These events, summarized in Table 1-3, feature student presentations exclusively.

TABLE 1-3 Opportunities for Undergraduates to Present Their Research**State and Regional Conferences**

- Georgia Undergraduate Research in Psychology Conference
- Southeastern Undergraduate Psychology Research Conference
- Arkansas Symposium for Psychology Students
- ILLOWA Undergraduate Psychology Conference
- Mid-America Undergraduate Psychology Research Conference
- Great Plains Students' Psychology Convention
- Joint Meeting of the Association for Psychological and Educational Research in Kansas and the Nebraska Psychological Society
- Michigan Undergraduate Psychology Paper Reading Conference
- Minnesota Undergraduate Psychology Conference
- Carolinias Psychology Conference
- Delaware Valley Undergraduate Research Conference
- Lehigh Valley Undergraduate Psychology Research Conference
- University of Winnipeg Undergraduate Psychology Research Conference

The journal *Teaching of Psychology* also carries a listing of undergraduate student conferences. For additional information you also might want to try the Society for the Teaching of Psychology Web page, which features an excellent list (<http://teachpsych.lemoyne.edu>).

If there is no student convention in your area, you can consider presenting a paper in one of the Psi Chi (National Honor Society in Psychology) sessions at a regional convention. One of the six regional association meetings held each year (Eastern, Midwestern, Rocky Mountain, Southeastern, Southwestern, and Western Psychological Associations) should be close enough to offer a potential forum for your research (see Table 1-4). In addition to the sessions held at these regional meetings, Psi Chi sponsors student paper sessions at the national meetings of the American Psychological Association and the American Psychological Society. Finally, if none of these options is a viable opportunity for you to present your research, then you should consider starting a paper-reading or poster presentation session on your own campus. Very successful annual events of this nature occur at many schools.

Although the opportunities for students to publish a journal article may be a bit more difficult to find than opportunities to present a paper at a convention, such opportunities do exist. For example, the *Journal of Psychological Inquiry* and the *Psi Chi Journal of Undergraduate Research* (see Table 1-5) are devoted to the publication of research conducted and

TABLE 1-4 Student Sessions Sponsored by Psi Chi (National Honor Society in Psychology)

Psi Chi routinely features paper and/or poster sessions at regional and national conferences.

Eastern Psychological Association

Rocky Mountain Psychological Association

Southeastern Psychological Association

Western Psychological Association

Midwestern Psychological Association

American Psychological Association

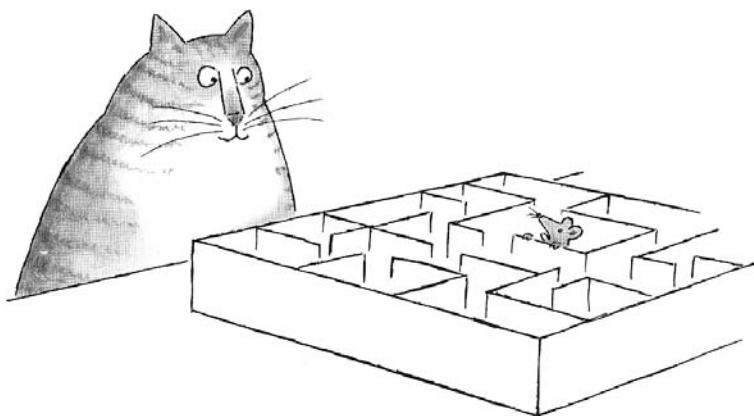
Southwestern Psychological Association

American Psychological Society

For information on Psi Chi and these sessions, contact:

Psi Chi National Office
P.O. Box 709
Chattanooga, TN 37043-0709
423-756-2044
psichi@psichi.org

The dates and locations of these conferences are routinely published in the *American Psychologist* (faculty members who belong to the APA receive copies of this journal) and *Teaching of Psychology*.



© The New Yorker Collection 1994 Samuel Gross from cartoonbank.com. All Rights Reserved.

"Well, you don't look like an experimental psychologist to me."

Attending a psychological convention will convince you that experimental psychologists are a most diverse group. Although you may not feel like an experimental psychologist now, completing a research project or two will change that situation.

reported by undergraduate students. If your faculty advisor has made a significant contribution to the design and conduct of your project, you may want to include him or her as a coauthor. *The Journal of Psychology and the Behavioral Sciences* is an annual journal that solicits manuscripts by students and faculty. Your faculty advisor will be able to suggest other journals to which you can submit your paper. Although undergraduates typically do not

TABLE 1-5 Publication Opportunities for Students

Several journals publish papers authored by students. Contact each journal to determine specific submission procedures.

1. *The Journal of Psychology and the Behavioral Sciences*

Professor John Brandi, Faculty Editor
Department of Psychology
Fairleigh Dickinson University
Madison, NJ 07904

2. *Modern Psychological Studies*

Department of Psychology
University of Tennessee at Chattanooga
Chattanooga, TN 37043-2598

3. *Journal of Undergraduate Studies*

Department of Psychology
Pace University
861 Bedford Road
Pleasantville, NY 10570

4. *Psi Chi Journal of Undergraduate Research*

Psi Chi National Office
P.O. Box 709
Chattanooga, TN 37043-0709
423-756-2044

5. *Journal of Psychological Inquiry*

Dr. Susan R. Burns
Department of Psychology
Morningside College
Sioux City, IA 51106
Burns@morningside.edu
402-280-3193

publish by themselves in the professional journals, collaborative papers featuring student and faculty authors are not uncommon.

The main point of this discussion is to encourage you to take advantage of the opportunities to share your research results with others. Much of the research you will read about in this book was conducted by students and then presented at conventions or published in journals such as the *Psi Chi Journal of Undergraduate Research* and *Journal of Psychological Inquiry*. If they can do it, so can you! Once you are involved in the research process, you will quickly find that it is a highly stimulating endeavor that never seems to end. There are always new problems to investigate.

Finding a New Problem

As you consider the relation of your experimental results to past research and theory and share your results with others who give you feedback, new research ideas will present themselves (see Horvat & Davis, 1998). Why didn't the results turn out exactly as predicted? Did you fail to account for some factor or variable? What would happen if you manipulated this or that variable in a different manner? The more deeply you immerse yourself in a research area, the more questions and problems you will find to research. As you can see from Table 1-1, we have now come full circle to the start of a new research project.

Why Is the Research Methods Course Important?

When students are asked, "Why are you taking a course in research methods (or experimental psychology)?" typical responses might be these:

"It's required for the major."
"I really don't know; I'll never conduct any research after this course."

As we go through this text, we hope to convince you that an understanding of research methods and data analysis can give you some real advantages in the field of psychology. Here are a few of those advantages:

- 1. Assisting You in Other Psychology Classes.** Because psychology's knowledge base rests on a foundation of research, it makes sense that much of what you will cover in your other psychology classes will consist of research examples. The more completely you understand research methodology, the better you will be able to master the material in your other classes. Although this point might make sense to you for courses such as learning or perception, even courses such as personality and abnormal psychology are based on research.
- 2. Conducting an Original Research Project.** Frequently, the research methods course includes or is followed by conducting an original student research project. If you have such an opportunity, take advantage of it and then plan to present and publish your research findings (refer to our earlier discussion of the research process).
- 3. Conducting a Research Project After Graduation.** Your authors learned a long time ago that it is smart to "never say never." This caution also applies to you as students of psychology. Consider the following example: Several years ago, a very bright student took the research methods course with one of your authors. Although this student found the class sessions interesting and intellectually stimulating, she disliked the material and vowed that she would never think about conducting research in psychology after the class was over. How wrong she was—her first job following graduation was conducting research at the Medical College of Virginia! If your career plans even remotely relate to the field of psychology, then the chances are quite good that you may have to conduct some type of research project as part of your job. Clearly, a course in research methods will provide a good understanding of what you will need to do in such instances. Even students who go

into nonpsychological fields may use their training on the job to conduct research on nonpsychological topics.

4. Getting into Graduate School. There is no getting around the fact that psychology graduate admissions committees view a course in research methods or experimental psychology very positively (Keith-Spiegel & Wiederman, 2000; Landrum & Davis, 2006). Your having completed such a course tells the admissions committee that you have a good grasp of basic research methodology. Graduate programs in psychology value such knowledge. Having presented or published a research report is also rated very highly by graduate school admissions committees (Landrum & Davis, 2006; Thomas, Rewey, & Davis, 2002).

5. Becoming a Knowledgeable Consumer of Research. Our society is flooded with knowledge claims. Many of these claims deal with psychological research and phenomena, such as the claims that a particular type of diet will improve your disposition, that IQ tests are good (or bad), that scientific tests have proved that this brand of cola tastes best of all, or that this toothpaste fights cavities better than all the rest. How do you know which of these claims to believe?

If you understand the research on which these claims are based and the “facts” presented as supporting evidence (or that there is no supporting evidence), then you are in a position to make a more educated decision concerning such knowledge claims. The research methods course will give you the basic foundation from which you can make educated decisions concerning knowledge claims you encounter in your everyday life.

■ REVIEW SUMMARY

1. Psychologists use a variety of research methods to answer numerous questions about human and animal behavior.
2. The research process consists of several interrelated, sequential steps: finding the problem, doing a literature review, taking theoretical considerations into account, making a hypothesis, choosing a research design, conducting the experiment, doing the data analysis and making statistical decisions, reviewing decisions in terms of past research and theory, preparing the research report, sharing your results, and finding a new problem.
3. A **theory** is a formal statement of relations among variables in a particular research area, whereas a **hypothesis** attempts to state predicted relations among variables in a selected portion of a theory.
4. The **research** or **experimental hypothesis** is the experimenter’s predicted outcome of a to-be-conducted experiment.
5. The **research design** specifies how the experimenter will (a) select participants, (b) form groups, (c) control extraneous variables, and (d) gather data.
6. We encourage students to submit their research reports for presentation at professional society meetings and for publication in journals.
7. The research methods course can (a) assist you in understanding research in other courses, (b) prepare you to conduct research after graduation, (c) increase your chances of being accepted to graduate school, and (d) make you a knowledgeable consumer of the results of psychological research.

Check Your Progress

1. Briefly describe the steps involved in the research process.
2. “A formal statement of the relation among relevant variables in a research area” best describes
 - a. acquisition of knowledge by authority
 - b. a logical syllogism
 - c. acquisition of knowledge by experience
 - d. a theory
3. You believe that giving rats a dose of vitamin C will improve their learning ability. This statement represents your
 - a. theory
 - b. experimental design
 - c. problem
 - d. hypothesis
4. Which of the following is a method of sharing your results with the scientific community?
 - a. presenting a paper at an undergraduate research conference
 - b. presenting a paper or poster at a psychological convention
 - c. publishing an article in a professional journal
 - d. any of the above
5. What should you do after you find your research problem?
 - a. develop a hypothesis
 - b. review the relevant literature
 - c. design your research project
 - d. find a relevant theory
6. Which of the following is a method of sharing your research results with the scientific community?
 - a. presenting a paper at an undergraduate research conference
 - b. presenting a paper or poster at a psychological convention
 - c. publishing an article in a professional journal
 - d. any of the above
 - e. none of the above
7. Other than fulfilling a requirement, what are the reasons for taking a research methods or experimental psychology course? Describe them.

Key Terms

Theory, 5

Hypothesis, 6

Research or experimental

hypothesis, 6

Research design, 6

Looking Ahead

In this chapter we provided you with a general introduction to how psychologists gather data. Subsequent chapters will build on and expand this general introduction. In Chapter 2 we will examine how you can find a researchable problem. Once we have identified the sources of research problems, then we will discuss the formulation of a good research hypothesis.

Developing a Good Research Idea and Conducting an Ethical Project

CHAPTER
2

The Research Idea

- Characteristics of Good Research Ideas
- Sources of Research Ideas

Developing a Research Question

Surveying the Psychological Literature

The Need for Ethical Principles

APA Principles in the Conduct of Research with Humans

- Is Deception in Research Necessary?
- Informed Consent • Participants at Risk and Participants at Minimal Risk • Vulnerable Populations • The Debriefing Session

The Ethical Use of Animals in Psychological Research

The Institutional Review Board

The Experimenter's Responsibility

The Participants' Responsibility

The Researcher's Ethical Obligations Once the Research Is Completed

- Plagiarism
- Fabrication of Data
- Lying With Statistics
- Citing Your References Correctly

The Research Idea

Research idea

Identification of a gap in the knowledge base or an unanswered question in an area of interest.

The starting point for your project is a research idea or problem. You find a **research idea** when you identify a gap in the current knowledge base or an unanswered question that interests you. For example, in Chapter 1 we saw that Kim Driggers and Tasha Helms were interested in whether salary level was related to willingness to date. They found this information was not available and conducted research to correct this deficiency. In this section we will examine the characteristics of good research ideas and then explore several sources for research ideas.

Characteristics of Good Research Ideas

All possible solutions to a detective case are not equally likely. Likewise, not all research ideas are equally good. Good ideas have certain characteristics that set them apart from less acceptable ideas and problems.

Testable The most important characteristic of a good research idea is that it is testable. Can you imagine trying to conduct research on a phenomenon or topic that you cannot measure or test? This situation would be like trying to answer the old question "How many

angels can dance on the head of a pin?" For many years people have regarded this question as being unanswerable because we have no way to measure the behavior of an angel. Although you may be chuckling to yourself and saying, "I would never ask that kind of question," remember our caution in Chapter 1 about never saying never.

Suppose, for example, you became interested in cognitive processes in animals. Although humans can describe their thoughts, a research project designed to measure cognitive abilities in animals *directly* is doomed to failure before it starts. At present, the best one can do in this situation is to provide an indirect measure of animal cognition. Consider, for example, the research conducted by Irene Pepperberg on an African gray parrot named Alex. Pepperberg (1994) reported that Alex was able to acquire the concept of "blue key," and when asked, "How many blue keys?" Alex was able to examine a group of 10 to 14 items and answer how many blue keys were present. Certainly, Alex appears to have had excellent cognitive abilities. The point we are making is that although some research problems, such as the number of angels dancing on the head of a pin, may never be testable, others such as animal cognition may lend themselves to evaluation through indirect tests. Moreover, just because a problem is not currently testable does not mean it will always remain in that category. You may have to wait for technology to catch up with your ideas. For example, scientists proposed synapses and neurotransmitters in the nervous system long before they were directly seen and verified.

Likelihood of Success If you stop to think about it, each research project is a contest to unlock the secrets of nature. (If we already knew all of nature's secrets, there would be no need for research.) Given that our view of nature is not complete, we must try to arrange our research project to be as close to reality as possible (Medewar, 1979). The closer our project comes to approximating reality, the greater the likelihood of successfully unlocking some of the secrets of nature. Sometimes our view of nature is not very clear and our research does not work very well; consider the following example.

In the 1980s researchers claimed that a new chemical compound, denatonium saccharide, was the most bitter substance in existence. Because denatonium saccharide offered intriguing practical applications, such as being an additive for plastic telephone and computer cable coverings and garbage bags to discourage animal pests, one of your authors and several of his students began to conduct research on this noxious chemical. Our view of nature was that denatonium saccharide was incredibly bitter and that all creatures great and small would react to this substance in the same manner. To verify this prediction, we began testing the aversiveness of denatonium saccharide with a variety of animals, ranging from rats, grasshopper mice, and gerbils to prairie dogs. Test after test yielded the same results: Our test animals did not behave toward denatonium saccharide as being especially bitter or aversive (Davis, Grover, Erickson, Miller, & Bowman, 1987; Langley, Theis, Davis, Richard, & Grover, 1987). Following an experiment in which human participants rated denatonium saccharide as significantly more bitter than a comparable solution of quinine (Davis, Grover, & Erickson, 1987), our view changed. Denatonium saccharide is very bitter as far as humans are concerned; when testing animals, however, it does not have the same effect. Had we not changed our view of nature, we might still be conducting experiment after experiment wondering why our test animals were not behaving as they were "supposed" to.

Hence, in addition to testability, a second characteristic of the good research idea is that your chances for success are increased when your view of nature approximates reality as closely as possible.



Other than by direct trial-and-error investigation, how can we determine the relevant factors in our chosen research area?

Examining past research is your best bet. Those variables that were effective in previous studies are likely to be the ones that will work in your experiment. We will have more to say about that topic presently.

Sources of Research Ideas

There are two general sources of research ideas: nonsystematic and systematic. We will discuss each source in some detail.

Nonsystematic sources Sources for research ideas that present themselves in an unpredictable manner; a concerted attempt to locate researchable ideas has not been made.

Nonsystematic Sources **Nonsystematic sources** include those occurrences that give us the illusion that a research idea has dropped out of the sky. These sources are nonsystematic in that we have not made any concerted effort to locate researchable ideas; they present themselves to us in a somewhat unpredictable manner. Although we refer to these sources as nonsystematic, we are not implying that the researcher is unfamiliar with the research area. Good researchers are familiar with the published literature and previous research findings. We do not seem to generate meaningful research ideas in areas with which we are not familiar.

Among the major nonsystematic sources of research ideas are inspiration, serendipity, and everyday occurrences.

Inspiration Some research ideas may appear to be the product of a blind flash of genius; in the twinkle of an eye an inspired research idea is born. Perhaps the most famous example of inspiration in science is Albert Einstein's (Koestler, 1964). Ideas just seemed to pop into his mind, especially when he was sailing. Although such ideas just seem to appear, it is often the case that the researcher has been thinking about this research area for some time. We see only the end product, the idea, not the thinking that preceded its appearance.

Serendipity A situation in which one phenomenon is sought but something else is found.

Serendipity **Serendipity** refers to those situations where we look for one phenomenon but find another. Serendipity often serves as an excellent source for research ideas. According to B. F. Skinner (1961), "When you run onto something interesting, drop everything else and study it" (p. 81).

Consider the following scenario in which Skinner described his reaction to the malfunctioning of a pellet dispenser in an operant conditioning chamber (commonly called a Skinner box):

As you begin to complicate an apparatus, you necessarily invoke a fourth principle of scientific practice: Apparatuses sometimes break down. I had only to wait for the food magazine to jam to get an extinction curve. At first I treated this as a defect and hastened to remedy the difficulty. But eventually, of course, I deliberately disconnected the magazine. I can easily recall the excitement of that first complete extinction curve. I had made contact with Pavlov at last! . . . I am not saying that I would have not got around to extinction curves without a breakdown in the apparatus. . . . But it is no exaggeration to say that some of the most interesting and surprising results have turned up first because of similar accidents. (Skinner, 1961, p. 86)

His initial reaction was predictable; he saw the malfunction as a nuisance and wanted to eliminate it. However, with the eyes of an insightful researcher, he saw beyond his temporary frustration with the broken equipment to a more important possibility. By disconnecting the pellet dispenser he could study extinction. All of the subsequent research conducted on extinction and the various schedules of reinforcement that were developed indicate that Skinner capitalized on this chance happening. It was truly a serendipitous occurrence.

Everyday Occurrences You do not have to be working in a laboratory to come in contact with good research ideas. Frequently, our daily encounters provide some of the best possibilities for research. Another incident from the life of Skinner provides an excellent example of this source for a research project:

When we decided to have another child, my wife and I decided that it was time to apply a little labor-saving invention and design to the problems of the nursery. We began by going over the disheartening schedule of the young mother, step by step. We asked only one question: Is this practice important for the physical and psychological health of the baby? When it was not, we marked it for elimination. Then the "gadgeteering" began. . . . We tackled first the problem of warmth. The usual solution is to wrap the baby in half-a-dozen layers of cloth—shirt, nightdress, sheet, blankets. This is never completely successful. Why not, we thought, dispense with clothing altogether—except for the diaper, which serves another purpose—and warm the space in which the baby lives? This should be a simple technical problem in the modern home. Our solution is a closed compartment about as spacious as a standard crib [see Figure 2-1]. The walls are insulated, and one side, which can be raised like a window, is a large pane of safety glass. . . . Our baby daughter has now been living [in this apparatus] for eleven months. Her remarkable good health and happiness and my wife's welcome leisure have exceeded our most optimistic predictions. (Skinner, 1961, p. 420)

This description of the original Air Crib appeared in 1945. Subsequently, the Air Crib was commercially produced and several hundred infants were raised in them. Thus, an everyday problem led to an interesting, if not unusual, research project for Skinner.

It also is clear that your ability to see a potential research project as you go through your daily activities depends on some knowledge of your field of interest. Because Skinner was knowledgeable in the techniques of behavior control, he was able to conceptualize and develop the Air Crib. Because your authors are not well versed in archaeology, we do not see



FIGURE 2-1 Skinner's Air Crib.

Source: Skinner, B.F. (1961). *Cumulative Record* (Expanded Edition), p. 420. New York: Appleton-Century-Crofts.

possible sites for excavations and discoveries. Conversely, an archaeologist would find it difficult to propose meaningful research problems in psychology.

Systematic sources Thoroughly examined, carefully thought-out sources for research topics.

Systematic Sources Study and knowledge of a specific topic are the basis of **systematic sources** of research ideas. Research ideas developed from systematic sources tend to be carefully organized and logically thought out. The results of past research, theories, and classroom lectures are the most common examples of systematic sources for research ideas.

Past Research As you read the results of past research, you gradually form a picture of the knowledge that has accumulated in a research area. Perhaps this picture will highlight our lack of knowledge, such as the influence of clothing on eyewitness credibility. On the other hand, you may find that there are contradictory reports in the literature; one research project supports the occurrence of a particular phenomenon, whereas another reports cast doubt on its validity. Perhaps your research project will be the one that isolates the variable(s) responsible for these discrepant findings! Your consideration of past research may also indicate that a particular experiment has been conducted only once and is in need of replication, or you may find that a particular project has been conducted numerous times and needs not replication but new research. In each of these instances our review of past research prompts a research project.

One specific type of past research that can serve as an excellent source of research ideas deserves special mention. The *failure to replicate* a previous finding creates an intriguing situation for the psychological detective. What features of the initial research resulted in the occurrence of the phenomenon under investigation? What was different about the replication that caused the results to be different? Only continued research will answer these questions. As Sherlock Holmes would say, "Data! Data! Data! I can't make bricks without clay!" (Doyle, 1927, p. 318).



After careful experimentation, Judith Mayer decides talking to her husband is not like talking to a brick wall.

© 2006 William Haefell from cartoonbank.com. All Rights Reserved.

Theory As we noted in Chapter 1, the two main functions of a theory are to organize data and to guide further research. The guidance function of a theory provides an endless panorama of projects for researchers who take the time and trouble to master the theory and understand its implications.

Consider, for example, *social facilitation theory*. According to Robert Zajonc (1965), the presence of other people serves to arouse a person's performance. In turn, this increased arousal energizes the most dominant response. If you are a skilled pianist, then the presence of others at a recital will likely yield a stellar performance; playing well is the dominant response in this situation. If you are just learning to play the piano, mistakes may be your dominant response and the presence of others may result in an embarrassing performance. Based on this theory, researchers have conducted over 300 social facilitation studies involving thousands of participants (Guerin, 1986). Theory truly does guide research!

Classroom Lectures Many excellent research projects are the result of a classroom lecture. Your instructor describes research in an area that sparks your interest, and ultimately this interest leads you to develop and conduct a research project. Although lectures may not seem to be strictly nonsystematic or strictly systematic sources of research ideas, we chose to include them as systematic sources because they often include an organized review of relevant literature.

For example, after hearing a lecture on conditioned taste aversion, one of our students, Susan Nash, became interested in the topic. Subsequently, she did a term paper and conducted her senior research project in this area. Moreover, several of the research projects she conducted as a graduate student dealt with the process by which animals acquire taste aversions. Yes, one lecture can contain a wealth of potential research ideas; it will pay rich dividends to pay close attention to lectures and keep careful notes when you see potential topics for research.

Developing a Research Question

Regardless of the source of your research idea, your first goal should be to turn your idea into a question. Having a question in mind will guide you during the activities summarized in the remainder of the chapter. With little effort, you can probably recall instances in which a question

Asking the right question is a key ingredient in a good research project.



"I still don't have all the answers, but I'm beginning to ask the right questions."

has guided your research efforts in the past. For example, you may have had a chemistry set when you were a child. Like most children with chemistry sets, you probably wondered to yourself, "What will happen if I mix this chemical with that chemical?" Again, like most children, you probably mixed the chemicals and found out! In the examples we cited previously, don't you imagine that B. F. Skinner thought to himself, "I wonder why the rat quit pressing the bar when the food magazine jammed?" or "I wonder if I can make taking care of a baby easier?" Those questions guided his research.

Likewise, we hope you remember times in classes when, after a particularly interesting point, you asked the instructor, "What would happen if they . . ." or "Has anyone ever tried to do that experiment this way?" Those are great examples of questions that could guide a research project for you. Once you have a question, all you need to do is find out how much psychologists already know about that question—you need to survey the literature.

Surveying the Psychological Literature

We have already mentioned the importance of being aware of past research several times in this chapter. However, looking up the past research in an area may seem to be an insurmountable task. Adair and Vohra (2003) aptly described this situation when they indicated that "[t]he knowledge explosion has created enormous difficulties for researchers to be aware of, access, and process the volume of new literature" (p. 15). How extensive is this knowledge explosion and how does it impact psychologists who are attempting to conduct research? Here's one example: In 1997, the number of published scientific abstracts was 3.7 million; this figure represents a substantial increase from the 550,000 published in 1957 and the 2.24 million published in 1977 (Kaser, 1998).

Given the vast amount of literature you may be dealing with, you need an organized strategy. Table 2-1 summarizes the procedures you should follow in reviewing the literature in a particular area.

The steps are as follows:

1. **Selection of Index Terms.** The key to conducting a search of the literature is to begin with terms that will help you access the relevant articles in your chosen area. You need to determine the psychological terms for the concepts you included in your research question. Psychologists may use different terms than the words you used in your question.

TABLE 2-1 Steps in Conducting a Search of the Literature

1. **Selection of Index Terms.** Select relevant terms for your area of interest from the *Thesaurus of Psychological Index Terms*.
2. **Computerized Search of the Literature.** Use the selected index terms to access a computerized database, such as PsycINFO.
3. **Obtaining Relevant Publications.** Use a combination of reading and note taking, photocopying, interlibrary loan, and writing or e-mailing for reprints to obtain needed materials.
4. **Integrating the Results of the Literature Search.** Develop a plan that will facilitate the integration and usefulness of the results of the literature search.

An excellent starting place for selecting terms is the APA's *Thesaurus of Psychological Index Terms* (2007). The first edition of the *Thesaurus* was based on the 800 terms used in *Psychological Abstracts* prior to 1973. These terms were expanded to include those that describe interrelationships and related categories, which is how you can find your question's concepts. Each edition of the *Thesaurus* is updated to include new terms.

Let's return to the investigation of salary as a factor in dating, which we introduced via the Driggers and Helms (2000) study in Chapter 1. Our research assistant, Casey Hadsall, a student at Emporia State University in Emporia, Kansas, conducted the following search of the literature. She began her literature review with the *Thesaurus*. Figure 2-2 shows a page from the *Thesaurus* that contains relevant terms for the general topic of social dating. She selected "social dating" as her key term.

2. Computerized Search of the Literature. Once she selected her key term, her next step was to use it to access a database. Currently, most colleges and universities turn to the APA's PsycINFO to search the published literature. Because the scope and nature of the database services offered by the APA can vary widely, you should familiarize yourself with the services available at your college or university. (You can learn more about APA's computerized databases at www.apa.org.) For example, your school may not have online access to abstracts of all articles published from 1887 to the present. If your online database is not complete, you will need to consult *Psychological Abstracts* for abstracts of earlier journal articles. In short, you cannot be sure your search is complete unless you know what your database covers. It is important to make your review of the literature as thorough as possible in order to avoid repeating previous mistakes and needlessly replicating research that has already been conducted time and again. Your computer search provides the author(s), title, journal, volume and page numbers, and an abstract of each article. All you need to do is enter your index term(s) from the *Thesaurus* and let the computer do the rest. (An example of a PsycINFO printout that Casey Hadsall obtained appears in Figure 2-3, page 23.)

Once you have determined the number of publications in your chosen area, you can limit or expand your choices as you see fit. Because easy-to-follow instructions are provided, running PsycINFO should pose no problems for you. If you are interested in just the most recent literature, you might also ask your librarian about PsycFIRST or check online at <http://www.apa.org/psycinfo/psycfirst.html> to learn more about PsycFIRST. PsycFIRST maintains only the last three years of entries and is handy for a quick scan of the most current literature.

What About the Internet? The amount of information that is available on the Internet staggers the imagination. Can this vast storehouse be tapped for the literature review? The answer is yes, but just as the good detective carefully evaluates the credibility of each witness and each piece of information, you must be very cautious. Let's see what the Internet can contribute on the topic of the effects of salary on dating. First, we selected a search engine, such as Google, and asked it to search for "dating." This search yielded 43 million hits! A quick perusal indicates that the vast majority of these entries are not appropriate to our task and can be disregarded. For example, sites such as *Amazing Love*, *Starmatch*, *Date-a-Doc*, and *Free Thinkers Match* deal with dating services and are not in line with our goal of gathering information on the effects of salary on dating. Likewise, the three sites that deal with "radioactive dating" are not appropriate to our research topic. The Internet did not assist us in this instance. However, many reputable journals and scientific groups publish quality material on the Internet, and it *can* be an excellent tool for the researcher. Google has developed a new search engine to meet the need of exclusively accessing more scholarly references; the URL is <http://scholar.google.com>. You might want to give it a try.

Social Control	RELATIONSHIP SECTION	Social Interaction
<p>Social Control — (cont'd)</p> <p>B Social Processes 1967 R ↓ Emotional Control 1978 ↓ Social Influences 1967</p> <p>Social Dating 1973 PN 1280 SC 48150 UF Dating (Social) B Human Courtship 1973 Interpersonal Interaction 1967 R Acquaintance Rape 1991 Couples 1962 Friendship 1967 Male Female Relations 1988 Premarital Intercourse 1973 ↓ Relationship Termination 1997 Romance 1997</p> <p>Social Demonstrations 1973 PN 76 SC 48160 UF Demonstrations (Social) Picketing B Social Behavior 1967 R ↓ Collective Behavior 1967 ↓ Political Participation 1988 ↓ Social Movements 1967 Student Activism 1973</p> <p>Social Density 1978 PN 497 SC 48165 SN Number of animals or humans per given space unit. For specifically high density conditions use CROWDING. UF Density (Social) Population Density R Crowding 1978 Overpopulation 1973 Personal Space 1973 ↓ Population 1973 ↓ Social Environments 1973</p> <p>Social Deprivation 1973 PN 291 SC 48170 SN Limited access to society's resources due to poverty, neglect, social discrimination, or other disadvantage. For a lack of social contact use SOCIAL ISOLATION. Consider also CULTURAL DEPRIVATION. B Social Processes 1967 Stimulus Deprivation 1973 N ↓ Social Isolation 1967 R Cultural Deprivation 1973 Disadvantaged 1967 ↓ Homeless 1988</p> <p>Social Desirability 1967 PN 1417 SC 48180 UF Desirability (Social) B Social Influences 1967 R Need for Approval 1997</p> <p>Social Development Use Psychosocial Development</p> <p>Social Discrimination 1982 PN 1071 SC 48185 SN Prejudiced and differential treatment based on religion, sex, race, ethnicity, disability, or other personal characteristics rather than on the basis of merit. Use a more specific term if possible. UF Discrimination (Social) B Discrimination 1967 Social Issues 1991 N Age Discrimination 1994</p>	<p>Social Discrimination — (cont'd)</p> <p>N Disability Discrimination 1997 Employment Discrimination 1994 Race and Ethnic Discrimination 1994 Sex Discrimination 1979 R Affirmative Action 1966 ↓ Civil Rights 1978 Racial and Ethnic Relations 1982 Racism 1973 ↓ Social Integration 1982 Stereotyped Attitudes 1967 Stigma 1991</p> <p>Social Drinking 1973 PN 496 SC 48190 SN Consumption of alcoholic beverages in social settings. B Alcohol Drinking Patterns 1967 Social Behavior 1967</p> <p>Social Environments 1973 PN 2507 SC 48200 B Environment 1967 N ↓ Academic Environment 1973 ↓ Animal Environment 1967 ↓ Communities 1967 Home Environment 1973 Poverty Areas 1973 Rural Environments 1967 Suburban Environments 1987 Towns 1973 ↓ Urban Environments 1967 ↓ Working Conditions 1973 R Cultural Deprivation 1973 Social Density 1978</p> <p>Social Equality 1973 PN 851 SC 48210 UF Equality (Social) B Social Issues 1991 R Affirmative Action 1985 ↓ Civil Rights 1978 Equal Education 1978 ↓ Human Rights 1978 ↓ Justice 1973 Racial and Ethnic Relations 1982 ↓ Social Integration 1982</p> <p>Social Facilitation 1973 PN 443 SC 48220 UF Facilitation (Social) B Social Behavior 1967 R ↓ Social Influences 1967</p> <p>Social Groups 1973 PN 1160 SC 48230 UF Cadres Cliques Groups (Social) N Dyads 1973 Ingroup Outgroup 1997 Minority Groups 1967 Reference Groups 1994 R ↓ Social Networks 1994</p> <p>Social Identity 1988 PN 1980 SC 48235 SN An aspect of self image based on in-group preference or ethnocentrism and a perception of belonging to a social or cultural group. N Professional Identity 1991 R Ethnic Identity 1979 Ethnocentrism 1973</p>	<p>Social Identity — (cont'd)</p> <p>R Ingroup Outgroup 1997 Minority Groups 1967 Reference Groups 1994 ↓ Self Concept 1967</p> <p>Social Immobility Use Social Mobility</p> <p>Social Influences 1967 PN 5700 SC 48250 UF Influences (Social) Coercion 1994 Criticism 1973 Enabling 1997 Ethnic Values 1973 ↓ Interpersonal Influences 1967 ↓ Power 1967 ↓ Prejudice 1967 Propaganda 1973 Social Approval 1967 Social Desirability 1967 Social Norms 1985 Social Values 1973 Superstitions 1973 Taboos 1973 R Authority 1967 ↓ Ethics 1967 Mentor 1985 Popularity 1988 Psychosocial Factors 1988 Reference Groups 1994 Role Models 1982 ↓ Social Behavior 1967 Social Change 1967 Social Comparison 1985 Social Control 1968 Social Facilitation 1973 ↓ Social Movements 1967 ↓ Social Reinforcement 1967</p> <p>Social Integration 1982 PN 1110 SC 48258 SN Process of uniting diverse groups (e.g., racial, ethnic, religious, or disabled) of a society or organization. In 1982, this term was created to replace the discontinued term RACIAL INTEGRATION. In 2000, RACIAL INTEGRATION was stripped from all records and replaced with SOCIAL INTEGRATION. UF Desegregation Integration (Racial) Racial Integration Segregation (Racial) B Social Issues 1991 Social Processes 1967 N School Integration 1982 R ↓ Activist Movements 1973 ↓ Civil Rights 1978 ↓ Mainstreaming 1991 Racial and Ethnic Relations 1982 ↓ Social Discrimination 1962 Social Equality 1973</p> <p>Social Interaction 1967 PN 7584 SC 48260 UF Interaction (Social) B Social Behavior 1967 N Encouragement 1973 ↓ Interpersonal Interaction 1967 Nonviolence 1991 Peace 1989 Physical Contact 1982 Victimization 1973 R ↓ Aggressive Behavior 1967</p>
		242

FIGURE 2-2 A Page from the *Thesaurus of Psychological Index Terms*.

Source: Copyright © 2004. Reprinted with permission of the American Psychological Association. All rights reserved.

PsycINFO

Journal of Youth & Adolescence, 1986 Dec Vol 15(6) 487-496

Characteristics of married and unmarried adolescent mothers and their partners.
Lamb, Michael E.; Elster, Arthur B.; Peters, Laura J.; Kahn, James S.; et al (U Utah, Salt Lake City, USA)

Summary: Studied a middle-class, urban-clinic sample of 275 mostly Caucasian adolescent mothers (aged 12-29 yrs) and their partners (aged 14-36 yrs). Couples married at the time of conception (n = 22) enjoyed more positive responses from prospective grandparents and earned more than couples not married at the time of conception, but they were more likely to be high-school dropouts, suggesting limits in their lifetime earning capacities and less likelihood in identifying one another as sources of emotional support. Couples who married between conception and delivery (n = 110) reported that prospective grandparents responded less favorably to the pregnancy, but while their current salaries were lower, they were more likely to be continuing with their education. Those married after conception had fewer antisocial and conduct disorders than Ss who chose to continue in a dating relationship. Overall, the couples who married after conception appeared to face less severe problems than either initially-married couples or steady daters. ((c) 1999 APA/PsycINFO, all rights reserved)

Key Phrase(s) parental approval & salary level & educational status, 12-19 yr old unmarried vs married before conception vs delivery, mothers & their 14-36 yr old partners

Maj Subject(s) Adolescent Fathers, Adolescent Mothers, Marital Status

Min Subject(s) Adolescence, Adulthood, Childhood, Couples, Educational Background, Family Relations, Income Level, Parental Attitudes, Unwed Mothers

Class. Code 2840

Classification Psychosocial & Personality Development

Population Human, Male, Female

Age Childhood (birth-12 yrs), Adolescence (13-17 yrs), Adulthood (18 yrs & older)

Language English

ISSN 0047-2891

Document Type Journal Article

Form/Content Type Empirical Study

FIGURE 2-3 A Printout from PsycINFO.

Source: Reprinted with permission of the American Psychological Association, publisher of the PsycINFO database, all rights reserved.

You must, however, be very careful when selecting information from the Internet to use in your literature review. What criteria should we use to evaluate sites? At present, although there are no universally accepted guidelines for evaluating Internet sites, we have found the following criteria to be useful (see Table 2-2).

3. Obtaining Relevant Publications. Once you have assembled a listing of books and journal articles relevant to your area of interest, it is time to acquire copies. There are several techniques for going about this task; you will probably use a combination of these strategies. First, find out what is available in your own library and plan to spend time reading this material and making photocopies for later use.

TABLE 2-2**Evaluating Internet Resources**

The WWW is a self-publishing medium in which anyone with the appropriate computer and software and an Internet connection can create and disperse information. Web pages should be evaluated carefully using some of the same criteria used to evaluate a print source. Be aware, however, that Web pages present additional evaluation challenges.

Authority	Who is the author? Is the author's name clearly visible? Is there a link to the author's e-mail address? What are the author's qualifications, education, and occupation? Has the author written articles or books other than Web pages? Check journal databases, such as PsycINFO. Is there a link to additional information about the author? A personal home page? Is the source peer-reviewed or edited? If so, by whom? Does the author belong to an organization? If the page is authored by an organization, what additional information about that organization is available? Check the domain name of the URL (.gov, .edu, .com, .org, .net, .mil).
Accuracy	Are there clues to tell you that the information on the page is true? Does the author list sources? Is there a bibliography on the page? Can the information be verified elsewhere—perhaps in a print source? Are there obvious errors (spelling, grammar, etc.)?
Objectivity	Does the page reflect a particular bias or viewpoint? Does the author use inflammatory or provocative language? What is the purpose of the page? To inform? To sell or market a product or an idea? To entertain? To persuade? Why was this page written and for whom? (The Web can be a "soapbox" for organizations or people.) Is there advertising on the page? If there is advertising, can it be differentiated from the informational content?
Currency	When was the page first published and last updated? Are the facts on the page current? When was the information gathered? Are the links current?
Coverage	Is the page a complete document or an abstract/summary? Does the author adequately cover the topic? What time period is covered? Are there links to additional coverage? Does the page contain information that is pertinent to your research topic? How can you use this information?

TABLE 2-2 (Continued)

Navigation/Design	Is the page easy to read? Does the background interfere with the page's content? Is the color "easy on the eyes"? Is the material presented in an orderly format? Do the graphics and images add to the presentation of the page? Are links appropriate to the topic of the page? Do they work? Is there a link at the bottom of the page to go back to the top? Is there a link on each supporting page to go back to the main page? Is there a comment link at the bottom of the main page? Do you need special software to view the information? Is there a fee to obtain the information?
--------------------------	---

Source: Binghamton University (<http://library.lib.binghamton.edu/search/evaluation.html>).

There are two options for obtaining items that are not available in your library. First, you can order them through the *interlibrary loan* department of your library. Most interlibrary loan services are reasonably fast and efficient, but there may be a small fee for providing copies of journal articles. Second, to obtain journal articles, you can write directly to the author to request a reprint (authors typically receive a supply of copies of each article they publish).

Writing for a reprint has an added advantage: It allows you to request additional, related articles published by the author. With the advent of e-mail and the Internet, the task of communicating with researchers in your area of interest has become fast and convenient. Most authors are pleased to comply with such requests, and there is no charge for reprints. Your literature search is likely to uncover publications that are several years, even decades, old. Because authors receive a large number of reprints when they publish an article, the chances are good they will have copies of even their older articles; don't hesitate to ask. Concerned about addresses for authors? Your computer search lists the institutional affiliation of the author(s); therefore, you should be able to look up the institutional address in any number of the guidebooks to colleges and universities that are in your library. If you have only the author's name, then consult your faculty members to see whether they have a membership directory for the APA or the Association for Psychological Science (APS). There is a good chance that an author you are trying to locate will be a member of one or both associations. If so, you can get the institutional and e-mail addresses from a current membership directory. Another alternative is to conduct a Google search; you might be amazed how well this technique works.

- 4. Integrating the Results of the Literature Search.** Once you have assembled the journal articles, book chapters, and books relevant to your research area, you will need to make sense of this material. This task can be formidable indeed; it will help if you have a plan.



© The New Yorker Collection 1998 Arnie Levin from cartoonbank.com.
All Rights Reserved.

Our students have found the following procedure to be quite effective. We hope you also find it useful in organizing the results of your literature search. As you read each article, keep good but succinct notes on each of the following aspects. Because most journal articles appear in the following sequence, your task will not be difficult. We will have more to say about how you actually write the results of your research in Chapter 14. For now, you need to master the task of summarizing what others have already written and published.

Reference Information. List the complete citation (in APA format; see Chapter 14) for the article you are abstracting. This information will facilitate completing the reference section of your research report.

Introduction. Why did the researchers conduct this experiment? What theory does this research seek to support?

Method. Use the following sections to describe how the project was conducted:

Participants. Describe the participants of the experiment. List such specifics as species, number, age, and sex.

Apparatus. Describe the equipment used by the researcher. Note any deviations from the standard apparatus, as well as any unusual features.

Procedure. Describe the conditions under which the participants were tested.

Results. Which statistical tests did the author use? What were the results of these statistical tests?

Discussion and Evaluation. What conclusions did the author reach? How do these conclusions relate to theory and past research? Describe any criticisms of the research that occurred to you as you read the article. (See the guidelines for critiquing research publications in Appendix XX.)

Once you have completed taking notes on these sections, you should condense them so all this information will fit on one side of a sheet of paper. As you prepare these single sheets, be sure always to use the same sequence of headings that you followed in making your notes. An example of a completed sheet appears in Figure 2-4.

Wann, D. L., & Dolan, T. J. (1994). Spectators' evaluations of rival and fellow fans. *The Psychological Record*, 44, 351-358.

Introduction

Even though "sports fans are biased in their evaluations and attributions concerning their team" (p. 351), no research has examined the evaluations spectators make of other spectators. Hence the present experiment was conducted to examine spectators' evaluations of home team and rival fans.

Method

Participants - One hundred three undergraduate psychology students received extra credit for participation.

Instruments - A questionnaire packet consisting of an information sheet, Sports Spectator Identification Scale, a five-paragraph scenario describing the behavior of a home team or rival spectator at an important basketball game, and several questions concerning the general behavior of the spectator described in the scenario was given to each participant.

Procedure - The questionnaire packet was completed after an informed consent document was completed and returned. One-half of the participants read the home team fan scenario, while the remainder of the participants read the rival team fan scenario. The determination of which scenario was read was determined randomly.

Results

Analysis of variance was used to analyze the data. The results of these analyses indicated the home team fan was rated more positively than the rival team fan by participants who were highly identified with the home team. This pattern of results was not shown by lesser identified fans. Of particular note was the finding that the highly identified participants did not rate the rival team fan more negatively than did the lesser identified participants; they just rated the home team fan positively.

Discussion and Evaluation

These results support the authors' initial predictions that participants would give more positive evaluations of fans rooting for the same team and more negative evaluations of fans rooting for a different team. Wann and Dolan's prediction that such evaluations would be shown only by fans who were highly identified with their team also was supported. These predictions were seen to be in accord with social identity theory. The study appeared to be well conducted. The fact that the study was not conducted at a sporting event limits its applicability.

FIGURE 2-4 Sample of a Summary of a Journal Article.

Source: Wann, D. L. & Dolan, T. J. (1994). "Spectators' evaluations of rival and fellow fans." *The Psychological Record*, 44, 351-358.

Because you put the same type of information in the same location on each page, it is easy to make comparisons of the various aspects and details between and among research projects. Additionally, using a single sheet for each reference allows you to arrange and re-arrange the references to suit your various needs: alphabetically to create a reference list, into separate stacks to represent positive and negative findings, and so forth.

As you have probably realized, the format we are suggesting is modeled after the accepted APA format for preparing papers (see Chapter 14). The experience you get in summarizing the results of your literature search will help you as you prepare a full-length APA-format paper.

Even though you may have the topic for your research clearly identified, you still are not ready to begin your project. As you will see in the next section, you and others will have to review your proposed project carefully in order to ensure that it meets approved ethical guidelines.

■ REVIEW SUMMARY

1. The starting point for the research project is the **research idea** or the identification of a gap in our current knowledge base.
2. Good research ideas should be testable and have a high likelihood of success (closely approximate nature).
3. **Nonsystematic sources** of research ideas give the impression that the idea simply appeared with little or no forethought. Inspiration, **serendipity** (looking for one phenomenon but finding another), and everyday occurrences are the major nonsystematic sources of research ideas.
4. Formal study and knowledge of a topic form the basis for **systematic sources** of research ideas. Past research and theories are the two main systematic sources of research ideas.
5. Developing a research question will help guide your research.
6. An organized approach is necessary to conduct a survey of the psychological literature. Such a search typically begins with the selection of index terms from the *Thesaurus of Psychological Index Terms*. Once you have selected these terms, you use them to access a computer database, such as PsycINFO or the *Psychological Abstracts*. After you have identified and made copies of the relevant literature, you must integrate it in a meaningful manner.

■ Check Your Progress

1. Matching
 1. research idea
 2. serendipity
 3. *Psychological Abstracts*
 4. PsycLIT/PsycINFO
 5. experimental hypothesis
 - A. look for one thing, find another
 - B. computer search of the literature
 - C. identification of a gap in the knowledge base
 - D. manual search of the literature
 - E. experimenter's predicted answer to research
2. Good research ideas must be _____ and have a high _____.

3. Distinguish between nonsystematic and systematic sources of research ideas. What are the specific ways each general source may be shown?
4. When we find something unexpected that is important in our research, what has occurred?
 - a. inspiration
 - b. everyday occurrences
 - c. serendipity
 - d. theory
5. Nonsystematic sources for research ideas are those
 - a. that follow a careful and logical examination of a particular area
 - b. that require reason and logic to bring the idea into a useable format
 - c. that fall outside the typical research areas of psychology
 - d. that come up unpredictably without an actual effort to find them
6. All of the following are examples of systematic sources of research ideas except
 - a. everyday occurrences
 - b. past research
 - c. theory
 - d. classroom lectures
7. The *Thesaurus* provides:
 - a. abstracts
 - b. operational definitions
 - c. index terms
 - d. experimental designs

The Need for Ethical Principles

Although it may appear that we are ready to begin our research project as soon as we have determined the nature of our research problem, this simply is not the case. We still have one major issue to deal with before we finalize our research design and start gathering data: We must consider the ethical nature of our research. The days of beating a confession out of a suspect with a rubber hose are gone for the detective. Likewise, psychologists must ask and answer such questions as "Are we putting our participants at risk?" "Is our experimental treatment harmful?" "Is the information we will gather from our experiment worth the potential risk and harm to the participants who are involved?" Despite older views to the contrary, science does not operate in a moral vacuum (Kimmel, 1988).

Under the 1974 National Health Research Act, the United States government demands assurance that federally funded projects have been reviewed and approved by a group of the proposing scientist's peers and that human participants have given informed consent concerning research participation. Since that time, seeking approval for a project from a group such as a Human Subjects Review Panel, an Animal Care and Utilization Committee, or an Institutional Review Board has become standard practice.

What has prompted such concern with ethics? Although we could cite many unethical practices, we will describe four instances that created major concern: the medical atrocities of World War II, the Tuskegee syphilis project, the Willowbrook hepatitis project, and Stanley Milgram's obedience studies of the 1960s.

During World War II, Nazi doctors conducted a lengthy series of experiments on civilian prisoners of war to determine the effects of various viruses, toxic agents, and drugs. The prisoners had no choice concerning whether they wanted to participate. After World War II many

of these doctors were tried in courts for the unethical practices they had inflicted on their unwilling participants. Many were found guilty and either hanged or given long prison sentences. The Nuremberg War Tribunal was responsible for developing a code of medical and research ethics (Sasson & Nelson, 1969). Among other things, the Nuremberg Code stressed consideration of the following ethical aspects of research:

1. Participants should consent to participate in research.
2. Participants should be fully informed of the nature of the research project.
3. Risks should be avoided whenever possible.
4. Participants should be protected against risks to the greatest extent possible.
5. Projects should be conducted by scientifically qualified personnel.
6. Participants have the right to discontinue participation at any time.

As we will see, the Nuremberg Code has had a major impact on the American Psychological Association (APA) code of ethics to which psychologists adhere.

In 1956 an outbreak of hepatitis at Willowbrook, a mental retardation facility, prompted doctors to begin *purposely* infecting newly admitted patients in order to study the disease's development under controlled conditions. Approximately 10% of the new admissions were infected, placed in a separate ward, and not given treatment (Beauchamp & Childress, 1979). Even though the patients' parents agreed to this procedure, a number of parents felt they had agreed under pressure. Although considerable information that improved the future treatment of hepatitis was gained from this project, it involved a major ethical problem. The ethical problem involved the use of the patients for research purposes not related to mental disability. It is difficult to envision a relation between mental disability and the investigation of hepatitis.

The Tuskegee syphilis study (Jones, 1981) began in 1932 and continued into the early 1970s. The purpose of the study was to observe the course of the syphilis disease in untreated individuals. To accomplish this goal, 399 African American men living in the area around Tuskegee, Alabama, who were infected with syphilis were recruited as participants. Doctors told the men they were being treated for syphilis by the U.S. Public Health Service (USPHS). The men were not told the real purpose of the study, nor did they ever receive any treatment for syphilis. Moreover, local physicians were told not to treat these men for syphilis, and the participants were told their USPHS treatment would be discontinued if they sought any sort of additional treatment for the disease. Although information about the course of syphilis may have been gained from this project, it was done at the expense of the unknowing participants, some of whom lost their lives.



Review the points of research concern from the Nuremberg Code. Which of these principles did the Tuskegee syphilis study violate? How did the study violate these principles?

It would appear that, with the exception of the principle concerning the conduct of the project by a qualified scientist, *all* of the principles of the Nuremberg Code were violated. The men did not consent to participate, nor were they fully informed about the project. The participants were not protected against risks, and they certainly did not appear to have the right to decline to participate.

In Milgram's (1963) obedience-to-authority study, each participant was assigned the role of "teacher" in what was portrayed as a "learning experiment." Another "participant" (actually an accomplice of the experimenter) played the role of "learner." The teacher's task was to teach the learner a series of words. The teacher was told the purpose of the experiment was to determine the effects of punishment on memory. Every time the learner, who was in an adjacent room, made a mistake, the teacher was to correct the learner by administering an electric shock. Even though no shocks were actually administered, the teacher (true participant) did not know it, and the learner always acted as if a shock had been administered.

Once the experimental session began, the learner purposely made mistakes so the teacher would have to administer shocks. With each mistake the experimenter instructed the teacher to increase the voltage level of the shock. As the mistakes (and voltage level) increased, the learner began to moan and complain and finally refused to answer any questions. The experimenter told the teacher to treat no response as a mistake and continue administering shocks. The session ended when the teacher (participant) reached the maximum level of shock or refused to administer any more shocks. Contrary to Milgram's initial prediction, the participants administered many shocks at what they assumed were *very high* voltage levels.

Although this research contributed to our understanding of obedience to authority, it was not without ethical problems. For example, rather than protecting the rights of the participants, the experimenter purposely made them feel discomfort and emotional distress by ordering them to continue administering the electrical shocks (Baumrind, 1964). Moreover, Milgram (1964) was not prepared for the amount of emotional upset his participants experienced. Even though this high level of discomfort was noted, Milgram made the decision to continue the research. To Milgram's credit, all participants received a debriefing session

What NOT to do
regarding research
ethics!



"We've got to draw a line on unethical behavior
and then get as close to that line as possible."

after they completed the experiment. During this session the researcher explained the purpose of the experiment and the “learner” talked with the real participant. Subsequently, all participants received a follow-up report describing the results of the experiment. Such debriefing and follow-up procedures were not mandatory in the early 1960s.

APA Principles in the Conduct of Research with Humans

Experiments such as the Tuskegee syphilis project and Milgram’s study have led to the development of ethical guidelines by the APA. The APA adopted and published the original code of ethics in 1973; it was revised in 1982 and again in 2002.

The current ethical standards that pertain to research and publication appear in Table 2-3.

Those standards dealing with (a) securing “informed consent” from the participants (Standard 8.02) and (b) using “deception” in research (Standard 8.07) have proved to be controversial.



Why do you think these principles have proved to be controversial in conducting psychological research?

Much psychological research, especially in the area of social psychology, has involved deception. Researchers believe that in many cases they can obtain honest and unbiased responses only when the participants are unaware of the true nature of the research. Hence, researchers find it difficult to give their participants a complete and accurate description of the experiment and to secure informed consent before the research is conducted.

Is Deception in Research Necessary?

Ideally, the researcher should be able to explain the purpose of a research project to the participants and enlist their cooperation. However, providing a *complete* explanation or description of the project may influence the participants’ responses. Consider, for example, the research conducted by Kim Driggers and Tasha Helms (2000) on salary and willingness to date (see Chapter 1). Had the participants in this study known they were to evaluate the effect of salary on dating, then their behavior could easily have been influenced by what they believed their responses were supposed to be. As we will see, it is very important in psychological research with human participants to try to ensure that the participants are not responding simply because (a) they have “figured out” the experiment and know how they are supposed to act or (b) they think the experimenter expects them to respond in a certain manner (see Rosenthal, 1966, 1985).

Hence, it is arguable that deception may be justified in some cases if our results are to be unbiased or uncontaminated by knowledge of the experiment and the expectancies that such knowledge may bring. If researchers use deception, how do they deal with obtaining informed consent for participation?

TABLE 2-3 APA Ethical Standard 8 for the Conduct of Research**8.01 Institutional Approval**

When institutional approval is required, psychologists provide accurate information about their research proposals and obtain approval prior to conducting the research. They conduct the research in accordance with the approved research protocol.

8.02 Informed Consent to Research

- (a) When obtaining informed consent as required in Standard 3.10, Informed Consent, psychologists inform participants about (1) the purpose of the research, expected duration, and procedures; (2) their right to decline to participate and to withdraw from the research once participation has begun; (3) the foreseeable consequences of declining or withdrawing; (4) reasonably foreseeable factors that may be expected to influence their willingness to participate such as potential risks, discomfort, or adverse effects; (5) any prospective research benefits; (6) limits of confidentiality; (7) incentives for participation; and (8) whom to contact for questions about the research and research participants' rights. They provide opportunity for the prospective participants to ask questions and receive answers.
- (b) Psychologists conducting intervention research involving the use of experimental treatments clarify to participants at the outset of the research (1) the experimental nature of the treatment; (2) the services that will or will not be available to the control group(s) if appropriate; (3) the means by which assignment to treatment and control groups will be made; (4) available treatment alternatives if an individual does not wish to participate in the research or wishes to withdraw once a study has begun; and (5) compensation for or monetary costs of participating including, if appropriate, whether reimbursement from the participant or a third-party payor will be sought.

8.03 Informed Consent for Recording Voices and Images in Research

Psychologists obtain informed consent from research participants prior to recording their voices or images for data collection unless (1) the research consists solely of naturalistic observations in public places, and it is not anticipated that the recording will be used in a manner that could cause personal identification or harm, or (2) the research design includes deception, and consent for the use of the recording is obtained during debriefing.

8.04 Client/Patient, Student, and Subordinate Research Participants

- (a) When psychologists conduct research with clients/patients, students, or subordinates as participants, psychologists take steps to protect the prospective participants from adverse consequences of declining or withdrawing from participation.
- (b) When research participation is a course requirement or an opportunity for extra credit, the prospective participant is given the choice of equitable alternative activities.

8.05 Dispensing with Informed Consent for Research

Psychologists may dispense with informed consent only (1) where research would not reasonably be assumed to create distress or harm and involves (a) the study of normal educational practices, curricula, or classroom management methods conducted in educational settings; (b) only anonymous questionnaires, naturalistic observations, or archival research for which disclosure of responses would not place participants at risk of criminal or civil liability or damage their financial standing, employability, or reputation, and confidentiality is protected; or (c) the study of factors related to job or organization effectiveness conducted in organizational settings for which there is no risk to participants' employability, and confidentiality is protected or (2) where otherwise permitted by law or federal or institutional regulations.

8.06 Offering Inducements for Research Participation

- (a) Psychologists make reasonable efforts to avoid offering excessive or inappropriate financial or other inducements for research participation when such inducements are likely to coerce participation.
- (b) When offering professional services as an inducement for research participation, psychologists clarify the nature of the services, as well as the risks, obligations, and limitations.

TABLE 2-3 (Continued)

8.07 Deception in Research

- (a) Psychologists do not conduct a study involving deception unless they have determined that the use of deceptive techniques is justified by the study's significant prospective scientific, educational, or applied value and that effective nondeceptive alternative procedures are not feasible.
- (b) Psychologists do not deceive prospective participants about research that is reasonably expected to cause physical pain or severe emotional distress.
- (c) Psychologists explain any deception that is an integral feature of the design and conduct of an experiment to participants as early as is feasible, preferably at the conclusion of their participation, but no later than at the conclusion of the data collection, and permit participants to withdraw their data. (See also Standard 8.08, Debriefing.)

8.08 Debriefing

- (a) Psychologists provide a prompt opportunity for participants to obtain appropriate information about the nature, results, and conclusions of the research, and they take reasonable steps to correct any misconceptions that participants may have of which the psychologists are aware.
- (b) If scientific or humane values justify delaying or withholding this information, psychologists take reasonable measures to reduce the risk of harm.
- (c) When psychologists become aware that research procedures have harmed a participant, they take reasonable steps to minimize the harm.

Source: American Psychological Association. (2002). Ethical principles of psychologist and code of conduct. *American Psychologist*, 57, 1060–1073. Reprinted with permission.

Informed Consent

Standard 8.02 indicates that the participants should give informed consent regarding their participation in a research project. This informed consent frequently takes the form of a statement concerning the research, which the participants sign before taking part in the research project. An example of an informed consent form appears in Figure 2-5.

As you can see, this document gives the participants a general description of the project in which they are going to participate and informs them that no penalties will be invoked if they choose not to participate. Also, it clearly states that the participants have the right to withdraw their participation at any time they desire.



Even though it may not be readily apparent, the process of informed consent has given researchers a new variable to evaluate. What is this variable?

The informed consent process itself and the information provided through this process have stimulated some interesting research. For example, Edward Burkley III, Shawn McFarland, Wendy Walker, and Jennifer Young (students at Southern Illinois University Edwardsville) evaluated the effect of having the “right to withdraw from the experiment” statement printed on

Informed Consent Document
<p>Read this consent form. If you have any questions ask the experimenter and he/she will answer the question.</p> <p>The Department of Psychology supports the practice of protection for human participants participating in research and related activities. The following information is provided so that you can decide whether you wish to participate in the present study. You should be aware that even if you agree to participate, you are free to withdraw at any time, and that if you do withdraw from the study, you will not be subjected to reprimand or any other form of reproach.</p> <p>In order to help determine the relationship between numerous personality characteristics, you are being asked to complete several questionnaires. It should not take more than 30 minutes for you to complete these materials. These questionnaires will be completed anonymously.</p> <p>"I have read the above statement and have been fully advised of the procedures to be used in this project. I have been given sufficient opportunity to ask any questions I had concerning the procedures and possible risks involved. I understand the potential risks involved and I assume them voluntarily. I likewise understand that I can withdraw from the study at any time without being subjected to reproach."</p>
_____ Date _____
_____ Participant and/or authorized representative

FIGURE 2-5 An Example of an Informed Consent Document.

the informed consent form. Two groups of students solved anagrams. One group of participants had the right to withdraw statement printed on their informed consent document, whereas the informed consent document for the second group did not contain this statement. Burkley et al. (2000) found that the students who had the statement printed on their form solved significantly *more* anagrams than students whose form did not contain this statement. Clearly, the informed consent process can have an impact on the results of an experiment. The influence of ethical procedures likely will offer researchers rich experimental opportunities for years to come.

Although the objectives of Standard 8.02 may not seem difficult to satisfy, one group of participants does create special problems—children. For example, if you want to conduct a project using first-graders, your research proposal initially must be approved by a college or university committee. (We will discuss Institutional Review Boards on pages 39–40.) After you have secured this approval, you need to contact the school system for approval. This approval may have to be granted at several levels: teacher, principal, superintendent, and school board. Each of these individuals will scrutinize your proposal thoroughly to safeguard the rights of the children. When you have successfully cleared these hurdles, you must secure permission from the parent or legal guardian of each child. The parent or legal guardian is the individual who will sign the informed consent form.

The more general the statement of the research project, the more room there is for the possible use of deception or for not completely explaining the project. For example, the

informed consent document shown in Figure 2-5 does not explicitly tell the participants that they are taking part in a study that will examine the relation between interpersonal flexibility, self-esteem, and fear of death (Hayes, Miller, & Davis, 1993).

By deceiving participants about the true purpose of an experiment or not providing complete information about the experiment, the experimenter may take the chance of placing some participants “at risk.” Who is to say that completing a particular survey or questionnaire will not provoke an intense emotional reaction in a given participant? Perhaps one of the participants in an investigation of child abuse was sexually molested as a preschooler and completing your questionnaire will reawaken those terrifying memories. You need to consider carefully all the possible reactions and evaluate their seriousness.

Participants at Risk and Participants at Minimal Risk

Participants at risk

Participants in an experiment that places them under some type of physical or emotional risk.

Participants at minimal risk

Participants in an experiment that does not place them under physical or emotional risk.

Participants at risk are participants who, by virtue of their participation in the research project, are placed under some emotional or physical risk. Certainly, the participants in the Tuskegee syphilis study and Milgram’s obedience studies were participants at risk. Securing informed consent from participants at risk is a mandatory condition.

Participants at minimal risk are participants who will experience no harmful effects through taking part in the research project. For example, Anastasia Gibson, Kristie Smith, and Aurora Torres (2000) observed the glancing behavior of customers at ATM machines (see Chapter 4). In this research the participants were at minimal risk; the recording of their glancing behaviors did not influence their physical or emotional well-being. Although desirable, it is not mandatory to secure informed consent from participants at minimal risk. In fact, Gibson et al. indicated, “Because this study was a systematic naturalistic observation, participants did not complete a consent form” (p. 149).

But what about those participants at risk who are participating in a study involving deception? How do we satisfy the ethical guidelines in such a case?

First, the researcher should tell the participants about all aspects of the research that are potentially harmful. If the researcher uses deception, the participants should not be deceived about some potentially harmful aspect of the research. For example, it would not be acceptable to tell participants they will be taking a vitamin tablet when the drug they are to consume is a hallucinogen. When participants are informed about all potentially harmful aspects of a research project, then they are able to give valid informed consent about participating.

The second procedure used to satisfy the ethical guidelines when using deception is to debrief the participants thoroughly after the experiment is completed. We turn to that topic after considering the needs of certain special populations.

Vulnerable Populations

What the researcher considers to be a risky or potentially harmful situation may depend on the type of participants chosen for a project. For example, experiments involving physical exertion, such as testing the effects of exercise on subsequent memory tasks, which pose no risk to healthy participants, may constitute a significant risk for persons in poor health. Likewise, the age of the participant may be a relevant consideration in determining potential emotional harm or distress in studies involving the selection of liked and disliked peers.

Researchers must exercise special care when the research participants may not be capable of fully understanding what participation in a project may entail. Such participant groups might include children; patients with physical or mental disorders; or persons with lower intelligence, low literacy, or English as a second language. In these instances the researcher must take special care to ensure that these individuals or their parents or legal guardians understand fully what is expected of them and any risks they may encounter in the research project. To understand the importance of taking these potentially vulnerable populations into account, imagine how you would feel if you were a college student in a foreign country where you didn't speak the language very well and you agreed to participate in a potentially hazardous research project without understanding the risks involved. In some instances professional organizations have developed specific guidelines for the treatment of such vulnerable populations. For example, the Society for Research in Child Development (SRCD) has developed a set of 16 guidelines for research with children. You can access these guidelines on the Internet at <http://www.srcd.org/about.html#standards>. Click on "SRCD Ethical Standards" under the "About SRCD" tab.

The Debriefing Session

The **debriefing session**, usually the final step in conducting the research project, involves explaining to the participants the nature and purpose of the project. Debriefing can be very important and should not be taken lightly by the researcher. Eliot Aronson and J. M. Carlsmith (1968) proposed several excellent guidelines for effective debriefing:

1. The researcher's integrity as a scientist must be conveyed to the participants. The researcher's personal belief in the scientific method is the foundation for explaining why deception was necessary and how it was employed in the research project.
2. If deception was used, the researcher should reassure the participants that it was not wrong or a reflection on their integrity or intelligence to feel that they have been tricked or fooled. Such feelings indicate that the deception used in the project was effective.
3. Because debriefing is usually the last step in the conduct of the research project, there may be a tendency to try to rush through it. This approach is not advisable. The participants have a lot of information to digest and understand; therefore, the debriefing session should progress slowly. The explanations should be clear and understandable.
4. The researcher should be sensitive to indications that the participants' discomfort is not being alleviated by the debriefing session and should strive to correct this situation. The goal of the debriefing session is to return the participants to the same (or close to the same) mental and emotional state they were in at the beginning of the project.
5. The researcher should repeat all guarantees of confidentiality and anonymity that were made at the beginning of the project. For such assurances to be seen as believable, the researcher must have clearly established his or her integrity as a scientist.
6. Do not try to satisfy debriefing requirements by saying that you will *send* an explanation and the results of the project at a later date. For maximum effectiveness, you should conduct the debriefing session *immediately* following the experimental session; there are no "easier" ways to satisfy this obligation.

Debriefing session The time at the conclusion of an experiment when its nature and purpose are explained to participants.

These guidelines, which define the ideal debriefing situation, should be followed as closely as possible. In most instances, the debriefing session will not be overly time-consuming. For example, Elvo Kuai-Long Sou, a student at Washington State University at Vancouver, and his faculty advisor, Lori Irving, surveyed the attitudes of U.S. and Macao college students toward mental illness. According to the researchers, "Once the participants finished the survey, each of them received a debriefing sheet that explained the purpose of the study and provided basic information regarding mental illness and how to seek help for a mental health issue" (Sou & Irving, 2002, p. 18).



Review the discussion of debriefing. What is the main goal of the debriefing session?

The main purpose of the debriefing session is to explain the nature of the experiment, remove or alleviate any undesirable consequences the participants may be experiencing, and generally try to return participants to the same state of mind they were in prior to the experiment. For example, Milgram was sensitive to the emotional state of his participants. After the research was complete, he sent the participants a five-page report that described the results and their meaning. The participants also completed a questionnaire; 84% indicated that they were glad they participated. In a one-year examination of 40 participants, a psychiatrist reported there was "no evidence . . . of any traumatic reactions" (Milgram, 1974, p. 197).

The debriefing session also can provide the experimenter with valuable feedback concerning the conduct of the experiment from the participants' viewpoint. Was the IV manipulation successful? If deception was used, was it successful? Were the instructions clear? The answers to these and other relevant questions can be obtained during the debriefing session.

The Ethical Use of Animals in Psychological Research

Up to this point our ethical considerations have dealt with research in which humans are the participants. Since Willard S. Small (1901) used the first rat maze at Clark University in 1900 (Goodwin, 2005), considerable psychological research has involved the use of animals. The use of animals in psychological research has created considerable controversy, debate, and even violence in recent years. Supporters of animal research point to the numerous accomplishments and scientific breakthroughs that are based on the results of animal studies (e.g., Kalat, 1992; Miller, 1985). For example, such medical procedures as blood transfusions, anesthesia, painkillers, antibiotics, insulin, vaccines, chemotherapy, CPR, coronary bypass surgery, and reconstructive surgery are based on animal studies. For over 100 years psychologists have used animals in studies of learning, psychopathology, brain and nervous system functioning, and physiology. It is undeniable that animal research has yielded numerous important and beneficial results.

Nevertheless, animal activists argue that the price humans have paid for such progress is too high. They point to problems with housing conditions, as well as the pain and suffering endured by research animals, and insist such treatment must stop (Miller & Williams, 1983; Reagan, 1983; Singer, 1975).

TABLE 2-4 APA Ethical Standard 8.09—Humane Care and Use of Animals in Research

- (a) Psychologists acquire, care for, use, and dispose of animals in compliance with current federal, state, and local laws and regulations, and with professional standards.
- (b) Psychologists trained in research methods and experienced in the care of laboratory animals supervise all procedures involving animals and are responsible for ensuring appropriate consideration of their comfort, health, and humane treatment.
- (c) Psychologists ensure that all individuals under their supervision who are using animals have received instruction in research methods and in the care, maintenance, and handling of the species being used, to the extent appropriate to their role. (See also Standard 2.05, Delegation of Work to Others.)
- (d) Psychologists make reasonable efforts to minimize the discomfort, infection, illness, and pain of animal subjects.
- (e) Psychologists use a procedure subjecting animals to pain, stress, or privation only when an alternative procedure is unavailable and the goal is justified by its prospective scientific, educational, or applied value.
- (f) Psychologists perform surgical procedures under appropriate anesthesia and follow techniques to avoid infection and minimize pain during and after surgery.
- (g) When it is appropriate that an animal's life be terminated, psychologists proceed rapidly, with an effort to minimize pain and in accordance with accepted procedures.

Source: American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060–1073. Reprinted with permission.

Although the debate between researchers and animal activists is likely to continue for some time, our present concern is with the ethical conduct of animal research. In addition to the concern over the ethics of human experiments, the 1960s witnessed an increase in concern for the ethical treatment of animals. For example, the Animal Welfare Act of 1966 was national legislation specifically passed to protect research animals. The APA has also established ethical standards for the use of animals in psychological research (APA, 2002). Table 2-4 presents these standards, which are designed to ensure humane treatment of animals used in research and for educational purposes. They have resulted in improved care and housing of animals.

Now that we have considered the ethical treatment of human and animal participants, let's examine the group that decides whether the proposed research should or should not be conducted. For the detective, the jury will make such decisions; in the case of the psychologist, the Institutional Review Board will determine the outcome.

The Institutional Review Board

At some institutions the review panel for the use of human participants may be the Human Subjects Review Panel; the Animal Care and Utilization Committee reviews proposals for animal research. At other institutions the **Institutional Review Board (IRB)** reviews both types of proposals. Although the exact name of the group may vary from institution to institution, its composition and functions are quite similar.

The typical IRB is composed of a cross-section of individuals. For example, if we examine a college or university IRB, we might find that it consists of faculty members from the fields of history, biology, education,

Institutional Review Board (IRB) The university committee that is responsible for determining whether a proposed research project conforms to accepted ethical standards.

psychology, and economics. Additionally, there will probably be one or two individuals who are not associated with the institution. A veterinarian must be a member of the panel that reviews animal research proposals.

The task of this group is not to decide on the scientific merits of the proposed research; the scientist proposing the research is the expert in that area. The IRB's responsibility is to examine the proposed procedures, any psychological tests or questionnaires that will be used, the informed consent document, plans for debriefing the participants, the use of painful procedures in animal research, procedures for disposing of animals humanely, and so forth. If participants are at risk, will they be made aware of this risk before the experiment is conducted? Do participants have the ability to terminate participation at any time? Will debriefing counteract the effects of deception if it has been used? Is it possible that a particular questionnaire or survey may evoke a strong emotional reaction? In short, the IRB serves to ensure that the experimenter treats research participants, whether they are humans or animals, according to the established ethical guidelines. If the IRB determines that a proposed project does not meet ethical standards, the IRB can request that changes be made to bring the project in line with standards.

When you conduct your own research, whether it is for your experimental psychology/research methods class or an independent study project, you undoubtedly will have to complete the paperwork and submit it to the IRB at your college or university. Do not be intimidated by this task. The IRB guidelines should be clear and straightforward. Prieto (2005) has developed a set of questions that will guide your thinking as you plan your project and prepare your IRB forms (see Table 2-5). We encourage you to refer to and use these guidelines when you prepare your IRB application.

The Experimenter's Responsibility

Ultimately the experimenter is the single individual who is accountable for the ethical conduct of the research project. In accepting this responsibility, the researcher carefully weighs the *benefits* and *costs* of the project and then decides whether to conduct it. The new knowledge produced by the project represents the benefit. What about costs? Such factors as time and expense clearly are costs and must be considered. Whether the research will place the participants at risk is another major cost.

Students doing research as part of the requirements of a course also are responsible for conducting their projects in an ethical manner. Applications to use human or animal participants must be submitted to and approved by the appropriate IRB. In most cases the supervising faculty member signs such submissions.

The Participants' Responsibility

Clearly, research participants have numerous, important rights. Let's turn the situation around.



Do participants also have ethical responsibilities? What should researchers be able to expect from their participants?

Korn (1988) indicated that the research participant has the following responsibilities:

1. ***Be on time for the research appointment.*** The researcher and other participants are inconvenienced when participants are late for an appointment.
2. ***Participants have the responsibility to listen carefully to the experimenter and ask questions in order to understand the research.*** Such attention and questions are important for the participants to give informed consent and “receive the maximum education benefit from participation” (p. 77).

TABLE 2-5 Questions to Ask and Answer as You Prepare Your IRB Forms

1. As you plan your project, ask yourself GENERAL QUESTIONS like:
 - What things could possibly go wrong with the way I plan to gather a sample, obtain informed consent, collect and store data, present and publish results? Take a worse case scenario approach.
 - What will participants’ experience of my project and data collection be like? Be empathic and put yourself in their shoes.
 - What are the general benefits of doing my study? How do these stack up against the potential risks or inconvenience I am placing on participants? Make sure the benefits of your project far outweigh risks to participants.
2. As you plan your project ask yourself more SPECIFIC QUESTIONS like:
 - Am I sure that any participant I might have in my research project can read and fully understand my informed consent materials for the study?
 - To what extent do my research materials solicit personally identifying features of participants so that there is a risk of them being known only through their research materials?
 - If data were accidentally lost or made public, what risk of embarrassment or loss of privacy would a participant face?
 - Have I checked with others to ascertain whether items in my research materials might be perceived as embarrassing, culturally offensive, or overly invasive to participants?
 - If I collect data on sensitive topics (e.g., sexual abuse), include protected populations (e.g., minors), or become aware of a participant’s potential harm to self or others (e.g., intent to commit suicide), do I know my obligations to act (if any) under institutional policies and state laws concerning the protection of minors or mentally ill persons?
 - If a participant starts to experience discomfort because of some of the questions on my instruments, have I made arrangements and appropriate referral services available to the person if the discomfort continues or is severe?
 - Can I make myself readily available to all participants during my research via an array of contact avenues (e.g., phone, fax, e-mail, fixed mailing address)?
 - Can I safely store my data in a locked location away from any eyes except those qualified co-investigators approved by IRB?
 - Can I make sure that incentives (e.g., extra credit, money, other valuables) are evenly and equitably available to all participants and that any incentive I use does not unduly entice a disadvantaged group or protected population to accept an unreasonable risk in participating in order to gain the incentive?

Source: Prieto, L. R. (2005). “The IRB and psychological research: A primer for students.” *Eye on Psi Chi*, 9(3), 24–25. Copyright © 2005 by Psi Chi, The National Honor Society in Psychology (www.psichi.org). Reprinted by permission. All rights reserved.

3. **Participants should take the research seriously and cooperate with the experimenter.** Regardless of the appearance of the research task, participants should take their role seriously and assume the researcher will provide an explanation later.
4. **When the study has been completed, participants share the responsibility for understanding what happened.** "Participants should ask questions that help to clarify the situation and enable them to learn more about the topic and method of research. Participants can be of great help by telling the researcher anything they think might be of interest" (p. 77).
5. **Participants have the responsibility for honoring the researcher's request that they not discuss the study with anyone else who might be a participant.** "Confidentiality works both ways" (p. 77). As part of the debriefing session, the researcher may ask the participants not to discuss the experiment with their friends. The participants are ethically responsible to honor this request; the disclosure of such information can invalidate the results of a research project (see Chapter 7).

Clearly, participants' responsibilities also must be factored into the ethical makeup of a research project. Unfortunately, participants' responsibilities usually receive the least amount of attention. Why? Many students become research participants because they are enrolled in introductory psychology, and completion of the course carries some type of research participation requirement. The use of sign-up sheets posted on a "research participant" bulletin board, or some other equally impersonal manner of contacting potential participants, makes a discussion of participant responsibilities most difficult. The solution appears to rest with the class instructor's obligation to explain and discuss these responsibilities and benefits with his or her students.

The Researcher's Ethical Obligations Once the Research Is Completed

The experimenter's ethical responsibilities do not end when the data are collected and the participants are debriefed. Experimenters are responsible for presenting the results of their research in an ethical manner. The two main problems encountered in this regard are plagiarism and fabrication of data.

Plagiarism

Plagiarism Using someone else's work without giving credit to the original source.

Plagiarism means using someone else's work without giving credit to the original author. Certainly, plagiarism is an obvious violation of the ethical standards of psychologists. Although it is difficult to believe that plagiarism occurs among established professionals, it does (Broad, 1980). Unfortunately, plagiarism is not an uncommon occurrence in colleges and universities. Although some students may view plagiarism as an easy way to complete an assignment or term paper, many students have told us that no one had ever explained plagiarism to them. The Department of Psychology at

Bishop's University (1994) has clearly summarized what you need to do to avoid plagiarism. The guidelines suggest the following:

1. Any part of your paper that contains the exact words of an author must appear in quotation marks, with the author's name, and the date of publication and page number(s) of the source.
2. You should not adapt material with only minor changes, such as combining sentences, omitting phrases, changing a few words, or inverting sentence order.
3. If what you have to say is substantially your own words, but you took the facts or ideas from a particular author, then omit the quotation marks and reference with a parenthetical citation like this: (Jones, 1949).
4. Always acknowledge secondary sources. (See p. 45.)
5. You must reference every statement of fact and every idea or opinion not your own unless the item is part of common knowledge.
6. Do not hand in for credit a paper that is the same as or similar to one you have handed in elsewhere.
7. It is permissible to ask someone to criticize a completed paper before you submit it and to bring to your attention errors in logic, grammar, punctuation, spelling, and expression. However, it is not permissible to have another person rewrite any portion of your paper or to have another person translate into English for you a paper that you have written in another language.
8. Keep rough notes and drafts of your work and photocopies of material not available in your college or university library.

In our experience, Guideline 2 is most problematic for students. If you reread this guideline, you will find that even paraphrasing an author's words is considered plagiarism. It is a good idea to read your sources thoroughly but to put them away when you write so you are not tempted to copy or paraphrase. For example, if you paraphrased the last sentence as "It is good to read your sources carefully but to put them away when you write so you will not be tempted to copy," you would be guilty of plagiarism.

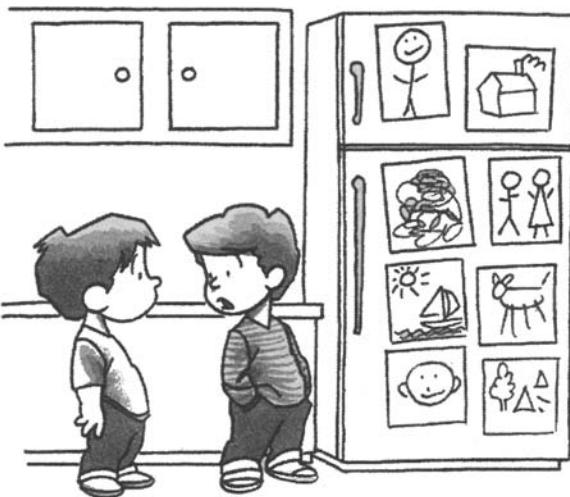
Why would a scientist engage in plagiarism? Although the specific reason(s) will vary from individual to individual, the pressure to publish one's research findings probably represents the single greatest cause (Mahoney, 1987). Because one's job security (tenure) or salary increases are often directly tied to the publication record, research and publication may become the central focus in the professional career. In haste to prepare articles and build an impressive publication record, it may seem easier and quicker to "borrow" a paragraph here and a paragraph there, especially if writing is not an easy or enjoyable task.

Fabrication of Data

Fabrication of data refers to those instances in which the experimenter either deliberately changes or alters data that were already gathered or simply makes up data to suit his or her needs. As with plagiarism, fabrication of data is an obvious ethical violation. Possibly the most famous instance of *apparent* data fabrication involved the famous British psychologist Cyril Burt (1883–1971). Burt's research on twins was influential in supporting the view that children's intelligence is largely inherited from their parents. His data went

Fabrication of data

Those instances where the experimenter either deliberately alters or creates research data.



© 2004 Reprinted courtesy Bunny Hoest and Parade Magazine.

"These days, it's publish or perish."

unchallenged for a number of years until researchers began to notice that he reported exactly the same results in different studies. Critics contended that it was very unlikely that exactly the same results would be produced in several different studies. Burt's supporters argued that Burt was elderly and likely forgot which data to present in which report. Unfortunately, scholars will not be able to resolve this debate because Burt's data were discarded shortly after his death. Researchers now agree that Burt's publications have little scientific value.



What would cause a scientist to fabricate data?

Again, we can point to the pressure to "publish or perish." Remember, we use synthetic statements in our research hypothesis; hence, there is always a chance that a project will not turn out as we predicted. When the data do not turn out as we had hoped, there may be a temptation to "fix" the data in order to bring them in line with our predictions and theory. Certainly, it is much easier (and quicker) to change a few numbers here and there than it is to redo an entire experiment. It does not seem to be a very great step to go from changing a few numbers to eliminating the experiment altogether and fabricating all the data.

The consequences of having false information in a body of scientific knowledge should be obvious to you. By falsifying a research finding, a psychologist actually encourages other researchers to conduct projects on a topic that is doomed to failure! The expenditures in time, effort, participants, and money can be enormous. And because much research is supported by federal grants, the costs to the taxpayers can be substantial. Because data form the cornerstone of research reports, ethical conduct is of vital importance.

Lying with Statistics

In addition to the sanctions against plagiarism and the fabrication of data, researchers have a responsibility to present their results in an unbiased manner. On first glance, this responsibility may seem rather straightforward and inconsequential. After all, results are results, aren't they? The answer to that question is yes and no. Yes, the basic results are not going to change; however, researchers exercise a great deal of individual choice and creativity in deciding what to present and how to present it. For example, a researcher could selectively choose to leave out important data and highlight other, possibly inconsequential, data. The choice of what to present and how to present it is so important in determining the conclusion your intended audience reaches that Darrell Huff (1954) published a book titled *How to Lie With Statistics* in which he described numerous "questionable" statistical practices designed to influence the consumer of this information. We will have more to say about the presentation of statistical results throughout the rest of this book. For now, you should be aware that researchers have the responsibility to present their results in an unbiased manner. Like a good detective, you need to be sensitive to those instances in which evidence may be misleading or missing.

Citing Your References Correctly

When other people read your research report and note the references you used to support the research idea, the hypothesis you are testing, and your interpretation of the data, they assume that you have actually read each reference that you cite. This assumption makes good sense; if you design a research project to test a theoretical prediction, then you should have read about that theory firsthand.

On the other hand, there may be occasions when you *read about* an interesting and relevant piece of research but you simply cannot acquire a copy of the original report. Perhaps your library does not carry the journal in which this article was published, and you cannot obtain the paper on interlibrary loan. Lacking the author's address, you are at a standstill. Perhaps the research has not been published. What should you do when you cannot obtain a copy of the original report? If the paper is from a journal to which you do not have access, it may be tempting to cite and list this reference *as if* you had actually read the original. We caution you against adopting this practice; it is the researcher's ethical responsibility to cite and list *only* those works that have been read. Instead, you should cite the *secondary source* you are using.

Don't despair; there is a way you can include those unobtainable references in your research report. Here's how it works. Assume that you read a 1999 article by Smith and Davis that *described and referenced* a research project that Brown conducted in 1984. You would like to cite the paper by Brown, but you are unable to obtain it. In this instance you cite the reference as follows:

Brown (as cited in Smith & Davis, 1999) found that . . .

In the reference section you would list only the Smith and Davis reference (the one you actually read). The appropriate way to cite and list references is covered in detail in Chapter 14. We encourage you to examine this material now. We agree that there are a lot of details to master when it comes to citing and listing references; please work hard on this skill. Sherlock Holmes's comment that "It has long been an axiom of mine that the little things are infinitely the most important" (Doyle, 1927, p. 194) is most appropriate in this context.

These, and other, APA ethical standards that are related to the presentation and publication of research data are shown in Table 2-6.

TABLE 2-6 APA Ethical Standards 8.10–8.15—Reporting and Publishing Your Research**8.10 Reporting Research Results**

- (a) Psychologists do not fabricate data.
- (b) If psychologists discover significant errors in their published data, they take reasonable steps to correct such errors in a correction, retraction, erratum, or other appropriate publication means.

8.11 Plagiarism

Psychologists do not present portions of another's work or data as their own, even if the other work or data source is cited occasionally.

8.12 Publication Credit

- (a) Psychologists take responsibility and credit, including authorship credit, only for work they have actually performed or to which they have substantially contributed.
- (b) Principal authorship and other publication credits accurately reflect the relative scientific or professional contributions of the individuals involved, regardless of their relative status. Mere possession of an institutional position, such as department chair, does not justify authorship credit. Minor contributions to the research or to the writing for publications are acknowledged appropriately, such as in footnotes or in an introductory statement.
- (c) Except under exceptional circumstances, a student is listed as principal author on any multiple-authored article that is substantially based on the student's doctoral dissertation. Faculty advisors discuss publication credit with students as early as feasible and throughout the research and publication process as appropriate.

8.13 Duplicate Publication of Data

Psychologists do not publish, as original data, data that have been previously published. This does not preclude republishing data when they are accompanied by proper acknowledgment.

8.14 Sharing Research Data for Verification

- (a) After research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through re-analysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release. This does not preclude psychologists from requiring that such individuals or groups be responsible for costs associated with the provision of such information.
- (b) Psychologists who request data from other psychologists to verify the substantive claims through re-analysis may use shared data only for the declared purpose. Requesting psychologists obtain prior written agreement for all other uses of the data.

8.15 Reviewers

Psychologists who review material submitted for presentation, publication, grant, or research proposal review respect the confidentiality of and the proprietary rights in such information of those who submitted it.

Source: American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060–1073. Reprinted with permission.

Now, we do not want to leave you with the impression that plagiarism, fabrication of data, and misrepresentation of results and references characterize the work of all or even many scientists; they do not. The vast majority of scientists enjoy designing and conducting their own research projects to see whether they can unlock nature's secrets. However, it takes only a few unethical individuals to give science a bad name. We hope you will not be among the ranks of the unethical.

■ REVIEW SUMMARY

1. It is the experimenter's responsibility to ensure that a research project is conducted in an ethical manner.
2. The World War II atrocities perpetrated by Nazi doctors, the Willowbrook hepatitis project, the Tuskegee syphilis project, and psychological research such as Milgram's obedience studies have increased ethical concerns and awareness.
3. The APA has developed a set of ethical standards for the conduct of research with human participants.
4. Although the researcher should try to avoid it, using deception may be justified in order to yield unbiased responses.
5. Informed consent is a signed statement indicating that participants understand the nature of the experiment and agree to participate in it.
6. **Participants at risk** are individuals whose participation in an experiment places them under some emotional and/or physical risk. **Participants at minimal risk** are individuals whose participation in an experiment results in no harmful consequences.
7. The experimenter explains the nature and purpose of the experiment to the participants during the **debriefing session**. Although the debriefing session is designed to counteract any negative effects of research participation, it also can provide the experimenter with valuable information about the research procedures that were employed.
8. Ethical standards also exist for the conduct of research using animals.
9. **The International Review Board (IRB)** is a group of the researcher's scientific and nonscientific peers who evaluate research proposals for adherence to ethical guidelines and procedures.
10. In addition to their rights and protection, participants also have several responsibilities.
11. Researchers have a responsibility to present their findings in an ethical manner. **Plagiarism** (the use of another's work without giving proper credit) and **fabrication of data** are the chief ethical infractions encountered in the presentation of research results.

■ Check Your Progress

1. Describe the three research projects that raised concern about the ethical conduct of research.
2. What aspects of the research endeavor are stressed by the Nuremberg Code?
3. In what types of research situations might the use of deception be justified? If deception is used, how can the need to provide informed consent be satisfied?

4. Participants in a research project who are placed under some emotional or physical risk are classified as
 - a. participants at risk
 - b. participants at minimal risk
 - c. volunteer participants
 - d. participants at no risk
5. The main purpose of the _____ is to explain the nature of an experiment and remove any undesirable consequences.
 - a. informed consent form
 - b. Institutional Review Board
 - c. Ethical Principles of Research With Human Participants
 - d. debriefing session
6. Describe the guidelines for the ethical use of animals in psychological research.
7. What is an IRB? Describe the composition of the typical IRB. What responsibility does an IRB have?
8. With regard to the conduct of research, what is the ethical responsibility of the experimenter?
9. Distinguish between plagiarism and fabrication of data. What sets of pressures are responsible for these unethical actions?

■ Key Terms

Research idea, 14	Participants at risk, 36	Plagiarism, 42
Nonsystematic sources, 16	Participants at minimal risk, 36	Fabrication of data, 43
Serendipity, 16	Debriefing session, 37	
Systematic sources, 18	Institutional Review Board (IRB), 39	

■ Looking Ahead

Now that we have discussed the sources of research problems, the conduct of an effective literature review, and the ethical considerations that must be taken into account before we conduct a research project, we will turn our attention to an actual research technique. In the next chapter you will learn about qualitative research methods.

Qualitative Research Methods

CHAPTER 3

General Overview

- Characteristics of Qualitative Research
- Data Analysis

Selected Examples of Qualitative Research Methods

- Naturalistic Observation • Ethnographic Inquiry
- Focus Groups • Case Studies • Grounded Theory
- Participatory Action Research

Most historians list 1879 as the year for the inception of modern psychology (Goodwin, 2005). The selection of that year to mark the birth of our discipline really is quite arbitrary; this was the year that Wilhelm Wundt and his students began conducting experimental, psychological research at the University of Leipzig in Germany. As we are sure you are already aware, the founding father of psychology wanted the new discipline to be a science; hence, heavy emphasis was placed on the conduct of experiments. In fact, the first major historical account of psychology was titled *A History of Experimental Psychology* (Boring, 1929). This scientific emphasis resulted in the establishment of psychology laboratories in both Europe and the United States. The conduct of laboratory-based experiments flourished and dominated the field of psychology until the middle of the 20th century.

As psychology grew and matured as a discipline, it became much more diverse both in its subject matter and in the people who entered the field as professionals. The diversity of subject matter made one fact crystal clear: The scientific or experimental method was not suited to investigate all the problems and questions that psychologists found themselves interested in studying. For example, psychologists interested in studying posttraumatic stress disorder certainly are not going to deliberately create a traumatic condition in order to study the development and course of this reaction.

Situations such as this one led researchers to develop a large number of *nonexperimental* research strategies. Although classified as nonexperimental, these methods do not necessarily produce inferior results. As you will see, the choice of a nonexperimental research strategy depends on the type of problem being investigated, where it is being investigated, and the nature of the data or information the researcher gathers. In this, the first of two chapters that deal with nonexperimental methods, we will examine a relatively new addition to the psychologist's research repertoire: qualitative research methods.

General Overview

Qualitative research

Research conducted in a natural setting that seeks to understand a complex human behavior by developing a complete narrative description of that behavior.

Qualitative research has its roots in the work of sociological and anthropological researchers, such as Franz Boas, Margaret Mead, and Bronislaw Malinowski. The German-born Boas really set the stage for qualitative research with his description of the Inuit people in the 1880s. His interest was not in changing the lives of these people, but rather in understanding their culture and how they lived. His was an approach that was interested in and looked at the entire society.

According to Featherston (2008), “One of the most important distinctions that sets qualitative research apart from more traditional types of research is that qualitative research is holistic in that researchers study phenomena in their entirety rather than narrowing the focus to specific defined variables” (p. 93). Similarly, Creswell (1984) indicated that qualitative research “is defined as an inquiry process of understanding a social or human problem, based on building a holistic picture, formed with words, reporting detailed views of informants, and conducted in a natural setting” (p. 2). Creswell’s definition clearly delineates the major characteristics of qualitative research.

Characteristics of Qualitative Research

Qualitative research is exemplified by the following characteristics:

1. It is an inquiry that seeks to understand social or human issues and problems. Because interactions with other people give meaning to things and events and because meanings change as we interact with various new people, qualitative research deals extensively with human interactions.
2. It seeks to develop an encompassing or holistic portrayal or picture of the problem or issue of interest. A qualitative project does not seek to reduce social situations or interactions to basic parts, elements, or values. Rather, the goal is to be able to put the reader of the research report into the social setting or situation that is the focus of the research.
3. It forms its report with words as opposed to statistical processes.
4. It uses the views of informants in the society as the basis for the report. In this regard, qualitative researchers do not begin their research with a preconceived, experimental hypothesis. They develop hypotheses and definitions as the research is ongoing.
5. It is conducted in a natural setting. In this regard it is important that qualitative researchers understand that when they undertake a project they have made a commitment to do extensive field research (Hatch, 2002). According to Featherston (2008), researchers “must spend enough time in the field to feel certain they are reporting what they claim” (p. 95).



Read through the five characteristics of qualitative research again. Based on these characteristics, describe the ways that a qualitative research report will differ from a more traditional experimental report prepared in accepted American Psychological Association (APA) format (see Chapter 14).

If you indicated that the qualitative research report generally will be more personal and will contain far less scientific terminology or jargon, you are definitely on the right track. If you also mentioned that the qualitative research report may not have a Results section that reports statistical measures and significance levels, you are right again.

Data Analysis

If the qualitative research report does not have a traditional Results section with numerical results and other data analyses, how do researchers evaluate the quality of their qualitative results? If qualitative researchers do not ask about validity (Do the data measure what they are supposed to measure?) and reliability (Do repeated measurements yield consistent results?), what types of questions do they ask? Rather than ask about validity and reliability, qualitative researchers are more likely to be interested in the *trustworthiness* of their data. Are the findings presented in this report worthy of attention? Hence, the issue has now become “how do researchers evaluate trustworthiness?” Guba and Lincoln (1994) proposed that trustworthiness can be judged by the following criteria:

1. **Confirmability.** Because the investigator who conducts qualitative research is the data recorder, there is the possibility that the report reflects the researcher’s subjective bias. In short, how can the researcher ensure that the report is accurate, unbiased, and can be confirmed by others? Qualitative researchers have several options that will satisfy this goal. First, they can have other researchers carefully peruse drafts of their report and point out inconsistencies, contradictions, and other possible instances of bias. Second, researchers should indicate the procedures they used to check and recheck the data throughout the study.
2. **Dependability.** Dependability is the extent to which the researcher believes that the same results would be produced if the study were replicated. Dependability in qualitative research is analogous to reliability in a research project that uses numerical data (i.e., a quantitative research project). The major difference is that in a qualitative research project the researcher must acknowledge and decide the nature of the context in which the research was conducted. This description is especially important if the context is volatile or in a state of change. Unlike a laboratory where the conditions can be uniform from one experiment to a replication of that experiment, qualitative research occurs in a natural setting that may change dramatically from one moment to the next.
3. **Credibility.** Whereas confirmability is concerned with removing bias from the qualitative researcher’s report, credibility is concerned with the accuracy of the identification and description of the subject of the study. If their focus is not accurate, it is possible for qualitative researchers to be unbiased and have excellent checks and rechecks on their data-gathering procedures and still not be studying the phenomenon they intended to study.

Confirmability

The extent to which the qualitative research report is accurate, unbiased, and can be confirmed by others.

Dependability

The extent to which the researcher believes that the same results would be produced if the study were replicated.

Credibility

The accuracy of the identification and description of the subject of the study.

If you indicated that you would have the participants (informants) read the report and judge its credibility, then you are absolutely correct. Because the qualitative research report is supposed to describe the phenomenon from the informants' point of view, they are the best judges of the credibility of the report.



Given what you know about qualitative research methods, how would you determine whether a qualitative research report is credible? Give this question some thought and propose a procedure(s) before reading further.

Transferability The extent to which the results of a qualitative research project can be generalized to other settings and groups.

4. **Transferability** (*generalizability*). Just like researchers who conduct projects in a controlled laboratory setting, qualitative researchers also are concerned with the generalizability or transferability of their findings to other settings and groups. To assist other researchers in making decisions about transferability, qualitative researchers must make their reports thorough and clear. Any assumptions they have made must be detailed. Then, because they have a thorough and clear report to follow, other researchers can make the judgment concerning the transferability of these findings to other settings and participants.

Selected Examples of Qualitative Research Methods

Naturalistic Observation

Naturalistic observation
Seeking answers to research questions by observing behavior in the real world.

Although this term is virtually synonymous with qualitative research, **naturalistic observation** is an identifiable method having several recommended guidelines. Among these guidelines are (a) determine the focus of the study, (b) plan the logistics, and (c) then plan the trustworthiness of the study (Lincoln & Guba, 1985). According to Featherston (2008), the goal of naturalistic inquiry "is to capture natural activity in natural settings" (pp. 96–97).

Ethnographic Inquiry

Ethnographic inquiry
Research in which the goal is to learn about a culture or some aspect of a culture from the perspective of the members of that culture.

Qualitative researchers conduct an **ethnographic inquiry** when their goal is to learn about a culture or some aspect of a culture from the perspective of the members of that culture (Glesne, 1999). In conducting an ethnographic project, the researcher typically uses participant observation. When using naturalistic observation, a researcher attempts to be as unobtrusive as possible; however, with **participant observation** the researcher becomes part of the group being studied. Yes, the participant observer is just like the detective who goes undercover to acquire information.

Participant observation
Research in which the researcher becomes part of the group being studied.

Participant observation is not limited to the study of human behavior; it can also be implemented in animal studies. Jane Goodall's study of chimpanzees in the Gombe Stream Reserve in Africa is a good example of the use of this technique with nonhumans. Goodall spent so much time observing the chimpanzees each day that

they finally treated her more as a member of the chimpanzee group than as an outside observer. Such acceptance facilitated her observations.

In the ethnographic approach you would spend a long time becoming immersed in the “culture” of the population being studied. We enclosed the word *culture* in quotation marks in the previous sentence because we do not refer to culture only in the cross-cultural sense of the word (e.g., different tribes, countries, continents). For example, you might wish to study an inner-city ghetto or a teenage gang using this method. As a researcher, you would become immersed in this culture for a long period of time and gather data through participant observation and interviewing.

Glesne (1999) differentiated between two types of participant observation depending on the degree of participation. The *observer as participant* refers to a researcher who primarily observes a situation but also interacts with the others involved in the study. She used the example of Peshkin’s (1986) semester-long study of a fundamentalist Christian school, in which the researchers predominantly sat in the back of the classroom and took notes, with little interaction. On the other hand, the *participant as observer* refers to a researcher who becomes a part of the culture by working and interacting extensively with the others. Glesne spent a year in St. Vincent, where she socialized, assisted in farm work, and even became an intermediary with governmental agricultural agencies.

There is a cost-benefit relation with these two approaches. The more immersed in a culture you become, the more you stand to learn about it, but at the same time, you stand to lose your objectivity about a culture as you become more immersed in it. To be a good participant observer, you probably would have to fit Dr. Watson’s description of Sherlock Holmes: “You really are an automaton—a calculating machine. There is something positively inhuman in you at times” (Doyle, 1927, p. 96).



The participant as observer technique has considerable appeal as a research procedure; the researcher is actually a part of the situation under study. What better way is there to acquire the desired information? Despite such appeal, this procedure has its drawbacks and weaknesses. What are they?

There are several potential drawbacks to the use of participant observation. First, an extended period of time may be necessary before the participant observer is accepted as a member of the group that is under study. Has such an extended period been budgeted into the overall design of the project in terms of both cost and time? Moreover, just being part of the situation does not guarantee that the observer will be accepted. If such acceptance is not granted, the amount of information that can be acquired is severely limited. If the observer is accepted and becomes part of the group, then a loss of objectivity may result. Thus, time, finances, acceptance, and objectivity may pose severe problems for participant observer research.

Focus Groups

A **focus group** consists of 7 to 10 participants who typically do not know each other but do have shared experiences or similar characteristics. These participants meet with a moderator for an hour to 1½ hours to discuss a particular topic. The moderator prepares 6 to 10 key questions that

Focus group Seven to 10 participants with shared experiences or similar characteristics who meet for 1 to 1½ hours to discuss a topic of common interest.

will help guide the discussion. The focus group is effective in determining how people feel or why they hold certain beliefs. Researchers frequently use this technique in conjunction with other qualitative methods.

Case Studies

Case studies Studies involving the intensive observation of a single participant over an extended period of time.

When we conduct **case studies**, we intensively observe and record the behavior of one or sometimes two participants over an extended period of time. Because there are no guidelines for conducting a case study, the procedures employed, behaviors observed, and reports produced may vary substantially.

Frequently, case studies are used in clinical settings to help formulate hypotheses and ideas for further research. For example, the French physician Paul Broca (1824–1880) reported the case study of a patient nicknamed “Tan” because he said nothing except the word *Tan* and an occasional obscenity when he was frustrated or annoyed (Howard, 1997). On the basis of his observations, Broca hypothesized that the center controlling the production of speech was located in the frontal lobe of the left hemisphere of the brain and that this area was damaged in the patient’s brain. An autopsy indicated that Broca’s hypothesis was correct. This case study provided the inspiration for numerous subsequent studies that yielded considerable information concerning the neurological system, which is responsible for the comprehension and production of speech. Other case studies might involve the observation of a rare animal in the wild or at a zoological park, a specific type of mental patient in a hospital, or a gifted child at school.

Although case studies often provide interesting data, their results may be applicable only to the individual participant under observation. In other words, when using the case study method, the researcher should not generalize beyond the participant who was studied. Additionally, because the researcher manipulates no variables, use of the case study method precludes establishing cause-and-effect relations. If your research goal, however, is to understand the behavior of one individual (such as Broca’s study of Tan), the case study may be the perfect research procedure. Your understanding of the behavior of one individual may lead to more general predictions.

Grounded Theory

Grounded theory A qualitative research approach that attempts to develop theories of understanding based on data from the real world.

Strauss and Corbin (1990) favored the **grounded-theory** approach to qualitative research. As with participant observation and the clinical perspective, the primary tools of discovery are interviews and observations. Grounded theory, however, goes beyond the descriptive and interpretive goals of these two approaches and is aimed at building theories. The ultimate goal of this approach is to derive theories that are grounded in (based on) reality. A grounded theory is one that is uncovered,

developed, and conditionally confirmed through collecting and making sense of data related to the issue at hand. The hope is that such theories will lead to a better understanding of the phenomenon of interest and to ideas for exerting some control over the phenomenon. Although grounded theory is designed to be a precise and rigorous process, creativity is also an important part of that process inasmuch as the researcher needs to ask innovative questions and come up with unique formulations of the data—to create new order out of the old” (Strauss & Corbin, 1990, p. 27). The grounded-theory approach is reminiscent of a detective’s attempt to build a theory about why certain types of crimes are committed. For

example, by interviewing a number of arsonists, you might be able to develop a theory about why arson takes place.

Strauss and Corbin (1990) did not advocate using grounded theory for all types of questions. In fact, they maintained that certain types of questions support certain types of research. For instance, if someone wanted to know whether one drug is more effective than another, then a grounded-theory study would not be appropriate. However, if someone wanted to know what it was like to be a participant in a drug study, then he or she might sensibly engage in a grounded-theory project or some other type of qualitative study.

The use of literature also differs in the grounded-theory approach. Strauss and Corbin (1990) recommended against knowing the research literature too well before using this approach because knowing the categories, classifications, and conclusions of previous researchers might constrain your creativity in finding new formulations. Instead, nontechnical materials such as letters, diaries, newspapers, biographies, and videotapes are essential to grounded-theory studies.

The heart of the grounded-theory approach occurs in its use of coding, which is analogous to data analysis in quantitative approaches. Three different types of coding are used in a more or less sequential manner (Strauss & Corbin, 1990). **Open coding** is much like the description goal of science. During open coding the researcher labels and categorizes the phenomena being studied. **Axial coding** involves finding links between categories and subcategories from open coding. The final step in the process, **selective coding**, entails identifying a core category and relating the subsidiary categories to this core. From this last type of coding, the grounded-theory researcher moves toward developing a model of **process** and a **transactional system**, which essentially tells the story of the outcome of the research. *Process* refers to a linking of actions and interactions that result in some outcome (see Figure 3-1 for a hypothetical process diagram). A transactional system is grounded theory's analytical method that allows an

Open coding The process of describing data through means such as examination, comparison, conceptualization, and categorization.

Axial coding The process of rearranging data after open coding so that new relations are formed between concepts.

Selective coding The process of selecting the main phenomenon (core category) around which all other phenomena (subsidiary categories) are grouped, arranging the groupings, studying the results, and rearranging where necessary.

Process The manner in which actions and interactions occur in a sequence or series.

Transactional system An analysis of how actions and interactions relate to their conditions and consequences.

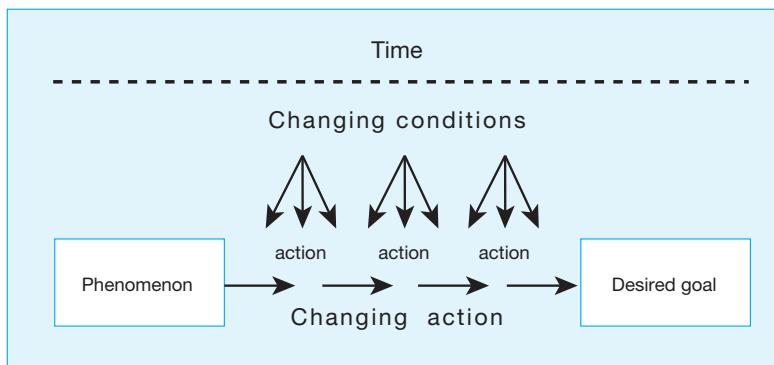


FIGURE 3-1 A Process Diagram from Grounded Theory Showing the Occurrence of Actions and Interactions in a Sequence.

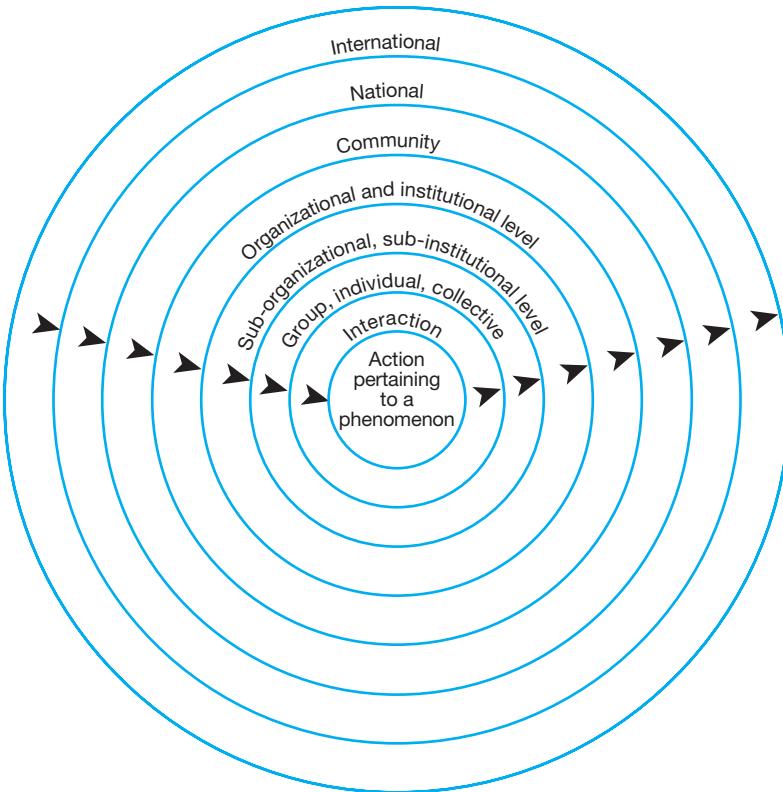


FIGURE 3-2 A Transactional System Conditional Matrix from Grounded Theory.

Conditional matrix

A diagram that helps the researcher consider the conditions and consequences related to the phenomenon under study.

examination of the interactions of different events. The transactional system is depicted in a **conditional matrix** such as that shown in Figure 3-2.

In the conditional matrix, the factors that are most pertinent to an event are shown at the interior; the least important factors, on the exterior. Once the coding is completed and a model of process or a transactional system has been developed, the grounded-theory procedure is complete. It then remains for the researcher to write a report about the findings (often a book rather than a journal article). Subsequent grounded-theory research projects tend to focus on different cultures. Comparisons between the original project and subsequent projects shed light on

whether the original project has generalizability to other cultures or whether it is specific to the culture studied.

Participatory Action Research

According to Fine et al. (2003), participatory action research (PAR) is a qualitative research approach “that assumes knowledge is rooted in social relations and most powerful when produced collaboratively through action” (p. 173). Researchers typically conduct a PAR project in a community; the goal of the project usually is to evaluate and understand the impact of some social program on the community. For example, Fine et al. were interested in evaluating the effects of taking high school and college courses on the attitudes and behaviors in inmates in

a prison (the community). The PAR project typically involves the participants—the prisoners in the Fine et al. research—as equal researchers (not subjects of study) in the project. Beyond this general framework, the PAR approach can use one or more specific methods to gather information; Fine et al. used several methods—both qualitative and quantitative. Their qualitative methods included interviews, focus groups, surveys, and narratives. Their quantitative method was a statistical analysis of the effects of attending college courses. The quantitative results showed that female inmates who took college courses in prison (a) had a very low return-to-custody rate, (b) were twice as likely *not* to be arrested for a new crime, and (c) were much less likely to violate their parole. The qualitative data showed that the prisoner-researchers who took college courses underwent a transformation. They were (a) more positive about themselves, (b) more concerned about the impression they made on others, (c) more self-controlled, and (d) more concerned about whether their behaviors would get them in trouble.

■ REVIEW SUMMARY

1. Because psychologists have a large number of diverse areas of research interest, they have developed a variety of research methods. Several of these methods are nonexperimental.
2. **Qualitative research** has its roots in the work of sociologists and anthropologists. This type of research is conducted in a natural setting and attempts to understand a social entity, such as a culture, or a human problem by developing a holistic narrative about the culture or problem.
3. Unlike experimental projects, qualitative research does not record numerical data and report the results of statistical analyses.
4. Qualitative researchers use **confirmability, dependability, credibility, and transferability** to evaluate the trustworthiness of a qualitative research report.
5. **Naturalistic observation** seeks answers to research questions by observing behavior in the real world.
6. The goal of an **ethnographic inquiry** is to learn about a culture or some aspect of a culture from the perspective of the members of the culture. To conduct an ethnographic inquiry, the researcher becomes a member of the group being studied (i.e., **participant observation**).
7. A **focus group** consists of 7 to 10 participants with shared experiences or similar characteristics who meet for 1 to 1½ hours to discuss a topic of common interest.
8. **Case studies** involve the intensive observation of a single participant over an extended period of time.
9. **Grounded theory** is a qualitative research approach that attempts to develop theories of understanding based on data from the real world.

■ Check Your Progress

1. Describe the characteristics of qualitative research.
2. Which of the following is *not* a criterion that researchers use to evaluate the trustworthiness of a qualitative research report?
 - a. credibility
 - b. validity
 - c. transferability
 - d. confirmability

3. The goal of _____ is to learn about a culture from the perspective of a member of that culture.
- naturalistic observation
 - an ethnographic inquiry
 - a focus group
 - participant observation
4. Participant observation is best associated with _____.
- naturalistic observation
 - a focus group
 - a case study
 - an ethnographic inquiry
5. The development of a theory is the ultimate goal in which approach?
- naturalistic observation
 - clinical perspective
 - focus group
 - grounded theory
6. Labeling and categorizing phenomena are associated with _____ coding.
- open
 - axial
 - selective
 - transactional
7. The linking of actions and interactions that result in an outcome is best associated with _____.
- transactional system
 - an ethnographic inquiry
 - participatory action research
 - ecological psychology
8. The research conducted by Fine et al. that evaluated the effects of taking high school and college courses on the attitudes and behaviors in inmates in a prison is an example of _____.
- grounded theory
 - ecological psychology
 - an ethnographic inquiry
 - participatory action research

■ Key Terms

Qualitative research, 50	Ethnographic inquiry, 52	Axial coding, 55
Confirmability, 51	Participant observation, 52	Selective coding, 55
Dependability, 51	Focus group, 53	Process, 55
Credibility, 51	Case studies, 54	Transactional system, 55
Transferability, 52	Grounded theory, 54	Conditional matrix, 56
Naturalistic observation, 52	Open coding, 55	

■ Looking Ahead

In the next chapter we will examine additional nonexperimental procedures that researchers use to gather information and data in situations where it may not be possible or desirable to conduct a scientific experiment. You will see that these nonexperimental methods are more focused than the qualitative methods we have just considered.

Nonexperimental Methods: Descriptive Methods, Correlational Studies, Ex Post Facto Studies, Surveys and Questionnaires, Sampling, and Basic Research Strategies

Descriptive Methods

- Archival and Previously Recorded Sources of Data
- Naturalistic Observation
- Choosing Behaviors and Recording Techniques

Correlational Studies

- The Nature of Correlations
- Correlational Research

Ex Post Facto Studies

Surveys, Questionnaires, Tests, and Inventories

- Surveys and Questionnaires
- Tests and Inventories

Sampling Considerations and Basic Research Strategies

- Sampling • Basic Research Strategies

Unlike the qualitative research techniques that we covered in Chapter 3, which also are nonexperimental procedures, the nonexperimental methods covered in this chapter will have a more narrowly focused research problem and will use statistical procedures to analyze the data.

Descriptive Methods

Because several of these methods do not involve the manipulation of any variables, they are called **descriptive research methods**. When we use descriptive methods, we can only speculate about causation that may be involved.

Archival and Previously Recorded Sources of Data

In some instances researchers may not gather their own data; they may answer their research question by using data recorded by other individuals for other purposes. For example,

Descriptive research methods Research methods that do not involve the manipulation of an independent variable.

public health and census data may be analyzed years later to answer questions about socioeconomic status, religion, or political party affiliation. In some instances the records and data you need to consult are stored in a central location. The Archives of the History of American Psychology were established at the University of Akron in 1974 for conducting research on the history of psychology. Although the letters, documents, and photographs contained in this collection were originally produced for numerous reasons, they are now used by researchers interested in answering questions about the history of our discipline.

You also can access several archival sources online. For example, the General Social Survey (GSS) has been conducted almost annually since 1972 by the National Opinion Research Center. The responses of more than 35,000 respondents to a wide range of socially relevant questions are available at <http://www.icpsr.umich.edu/gss/home.htm>. The number of research questions that can be addressed with this huge body of information seems endless. There is no charge for accessing this site; we encourage you to peruse it.

Not all sources of previously recorded data are stored in an archive or library or are available online for our use. There are numerous real-life sources. For example, Laura Marzola, a student at King's College (Wilkes-Barre, PA), and her faculty supervisor, Charles Brooks, examined back issues of the *Psi Chi Journal of Undergraduate Research* to determine which journals student researchers cited most frequently in their published articles (Marzola & Brooks, 2004). Similarly, Jennifer Salhany, a student at St. John's University (Staten Island, NY), and her faculty supervisor, Miguel Roig, used college and university handbooks and catalogs found on the Internet to determine the prevalence and nature of academic dishonesty policies (Salhany & Roig, 2004).

Moreover, if you are interested in differences in sexual preoccupation between men and women, you might choose to examine graffiti on the walls of public restrooms.



The use of archival and previously recorded data is certainly in line with being a psychological detective. You are putting together bits and pieces of data to answer research questions. Unfortunately, several problems are associated with this approach to gathering information. Consider conducting this type of research and see what problems you can discover.

Potential Problems There are several potential problems associated with using archival and previously recorded sources of data. First, unless you are dealing with the papers and documents of a few clearly identified individuals, you will not know exactly who left the data you are investigating. Not knowing the participants who make up your sample will make it difficult to understand and generalize your results. Consider the graffiti example. You may choose to record graffiti from restrooms on your campus. Who created the graffiti? Was it created by a representative sample of students? Although common sense may tell you that the sample of graffiti writers is not representative of students on your campus, let alone college students in general, you do not know. Your ability to make statements other than those that merely describe your data is severely limited.

Second, the participants may have been selective in *what* they chose to write. Clearly, this consideration may be important in our graffiti example, particularly if we are evaluating the presence of sexual comments. What others chose to record may drastically influence our

conclusions in other instances. For example, until recently what we knew about Wilhelm Wundt, the founder of scientific psychology, was provided by E. B. Titchener, who originally translated Wundt's books and papers from German into English. Unfortunately, Titchener chose to misrepresent Wundt on several occasions; hence, our impression of Wundt may be severely distorted (Goodwin, 2005). Fortunately, Wundt's original writings are still available for retranslation and examination. Even if his works are retranslated, we will still face the possible problem that Wundt may have omitted things from his writings which he did not want to share with others. Whenever archival or previously recorded sources are used, you cannot avoid this problem of *selective deposit*.

A third problem with this type of data concerns the survival of such records. In our study of graffiti it will be important to know something about the cleaning schedule for the restroom(s) we are observing. Are they scrubbed clean each day, or are graffiti allowed to accumulate? In this example the data in which you are interested will probably not have a very high survival rate. Printed materials may not fare much better. During the 1920s John Watson and his call to behaviorism made psychology immensely popular in the United States; the number of journal articles and books written during this period attests to its popularity. It was only recently, however, that researchers discovered a very popular magazine, *Psychology: Health, Happiness, and Success*, which was published from 1923 to 1938 (Benjamin, 1992). Why does mystery surround this once-popular magazine? The problem had to do with the type of paper on which the magazine was printed. The high acid content of the paper led to rapid disintegration of these magazines; hence, only a precious few have survived.

Comparisons with the Experimental Method Certainly, the researcher can gain valuable information from archival and prerecorded sources. However, we must be aware of the problems of a nonrepresentative sample, data that are purposely not recorded, and data that have been lost. A comparison of using this technique with conducting an experiment reveals other weaknesses in addition to these limitations. Because we examined data and documents that were produced at another time under potentially unknown circumstances, we are not able to exercise any control with regard to the gathering of these data. We are therefore unable to make any type of cause-and-effect statement; the best we can do is speculate about what might have occurred.

These concerns notwithstanding, this type of research can yield interesting and valuable results. For example, much of what you read in a history of psychology text is the product of archival research. In the next section we examine methods in which we observe the phenomenon of interest firsthand.

Naturalistic Observation

As we saw in Chapter 3, **naturalistic observation** involves seeking answers to research questions by observing behavior in the real world and is the hallmark of qualitative research. However, researchers also use naturalistic observation to collect numerical data and answer more focused research questions. For example, each spring, animal psychologists interested in the behavior of migrating sandhill cranes conceal themselves in camouflage blinds to observe the roosting behavior of these birds on the Platte River in central Nebraska. For another example, a researcher who is interested in the behavior of preschool children might go to a day-care center to observe the children and record their behavior.

Naturalistic observation Seeking answers to research questions by observing behavior in the real world.

The possibilities for conducting studies using naturalistic observation are limited only by our insight into a potential area of research. Regardless of the situation we choose, we have two goals in using naturalistic observation. The first goal should be obvious from the name of the technique: to describe behavior as it occurs in the natural setting without the artificiality of the laboratory. If the goal of research is to understand behavior in the real world, what better place to gather research data than in a natural setting? The second goal of naturalistic observation is to describe the variables that are present and the relations among them. Returning to our sandhill crane example, naturalistic observation may provide clues concerning why the birds migrate at particular times of the year and what factors determine the length of stay in a certain area.

In a naturalistic observation study, it is important that the researcher not interfere with or intervene in the behavior being studied. For example, in our study of preschoolers, the observer should be as inconspicuous as possible. For this reason, the use of one-way mirrors, which allow researchers to observe without being observed, is popular.



Why should the researcher be concealed or unobtrusive in a study using naturalistic observation?

The main reason the researcher must be unobtrusive in studies using naturalistic observation is to avoid influencing or changing the behavior of the participants being observed. The presence of an observer is not part of the natural setting for sandhill cranes or preschoolers; they may well behave differently in the presence of observers.

Reactance or reactivity effect The finding that participants respond differently when they know they are being observed.

The **reactance** or **reactivity effect** refers to the biasing of the participants' responses because they know they are being observed. Perhaps the most famous example of a reactivity effect occurred in a study conducted at the Western Electric Company's Hawthorne plant located on the boundary between Chicago and Cicero, Illinois, in the late 1930s (Roethlisberger & Dickson, 1939). The purpose of the research was to determine the effects of factors such as working hours and lighting on productivity. When researchers compared the productivity of the test participants to that of the general plant, an unusual finding emerged. The test participants produced at a higher rate, often under test conditions that were inferior to those normally experienced. For example, when the room lighting was reduced well below normal levels, productivity increased. What caused these individuals to produce at such a high rate? The answer was simple: Because these workers knew they were research participants and that they were being observed, their productivity increased. Thus, the knowledge that one is participating in an experiment and is being observed may result in dramatic changes in behavior. Because of the location of the original study, this reactivity phenomenon is often referred to as the **Hawthorne effect**. Having considered the general nature of naturalistic observation, we will now examine a specific observational project more closely.

Hawthorne effect
Another name for reactance or reactivity effect.

Anastasia Gibson and Kristie Smith, students at the University of Alabama in Huntsville, and their faculty advisor, Aurora Torres, conducted an interesting study

that used naturalistic observation. They wanted to examine the relation between the glancing behavior of people using an automated teller machine (ATM) and the proximity of other customers. On the basis of previous studies of social distance and personal space, they predicted that other potential customers would not invade the personal space of the person using the ATM and that glancing behavior would decrease as other customers approached the 4-ft. radius that defined the personal-space boundary. The researchers were “in inconspicuous locations where [they] could view the ATM” (Gibson, Smith, & Torres, 2000, p. 150). Contrary to their predictions, the results indicated that glancing behavior *increased* as proximity between the person using the ATM and other potential customers *decreased*.

As you may have surmised by now, the main drawback with the use of naturalistic observation is, once again, the inability to make cause-and-effect statements. Because we do not manipulate any variables when we use this technique, such conclusions are not possible.

Why use naturalistic observation if it does not allow us to make cause-and-effect statements? The first reason is quite straightforward: Naturalistic observation may be our only choice of research techniques to study a particular type of behavior. Psychologists who are interested in reactions to natural disasters, such as hurricanes, earthquakes, tornadoes, and fires, cannot ethically create such life-threatening situations just to study behavior; they must make their observations under naturally occurring conditions. Conducting an experiment in which researchers manipulate variables is only one of many legitimate techniques used to gather data.

A second reason for using naturalistic observation is as an adjunct to the experimental method. For example, you might use naturalistic observation before conducting an experiment to get an idea of the relevant variables involved in the situation. Once you have an idea about which variables are (and are not) important, you can conduct systematic, controlled studies of these variables in the laboratory setting. After you have conducted laboratory experiments, you may want to return to the natural setting to see whether the insights gained in the laboratory are indeed mirrored in real life. Hence, psychologists may use naturalistic observation before and after an experimental research project to acquire further information concerning relevant variables.

As with the other observation techniques, the ability to make cause-and-effect statements is the second problem with participant-as-observer research. Even though the participant observer may be close to the source of relevant information, no attempts are made to manipulate IVs or control extraneous variables.

Choosing Behaviors and Recording Techniques

It is one thing to say you are going to conduct an observational study, but it is another task actually to complete such a project. Just because the researcher does not manipulate the variables does not mean that a great deal of planning has not taken place. As is true with *any* research project, you must make several important decisions before beginning the project. Let's examine several of these decisions.

It seems simple enough to indicate that all behaviors will be observed; hence, everything of interest will be captured. Saying may be quite different from (and much easier than) doing. For example, observing and recording all behaviors may necessitate using video equipment, which may, in turn, make the observer identifiable. A participant observer with a video camera would probably not be especially effective. Had Anastasia Gibson and her colleagues (2000) used video equipment in their naturalistic study of glancing behavior at the ATM, their

obvious presence could have influenced the behavior of their participants. Hence, they remained as inconspicuous as possible when they observed the customers.

Time sampling Making observations at different time periods.

Instead of observing at exactly the same time each afternoon, the ATM researchers used a procedure known as time sampling. **Time sampling** involves making observations at different time periods in order to obtain a more representative sampling of the behavior of interest. The selection of time periods may be determined randomly or in a more systematic manner. Moreover, the use of time sampling may apply to the same or different participants. If you are observing a group of preschoolers, using the time-sampling technique will allow you to describe the behavior of interest over a wide range of times in the same children. However, using time sampling may purposely result in the observation of different participants and increase the generality of your observations. On the other hand, Gibson et al. did not make any nighttime observations; consequently, their results may not apply to *all* times of day. In fact, they indicated that “it would be interesting to see whether or not proximity or glancing behavior are influenced by the amount of sunlight at hand” (Gibson et al., 2000, p. 151).

It also is important to note that Gibson et al. (2000) observed customers at four different ATMs: Two machines were located outside (“one outside a mall and one in an urban shopping area,” p. 150), whereas “two indoor ATMs were inside discount department stores” (p. 150). Why did they make observations in these different locations? Had they limited their observations to one ATM in one location, they would not be able to generalize their results beyond

that one machine to other ATMs. By observing four different ATMs, they used a technique known as situation sampling. **Situation sampling** involves observing the same type of behavior in several different situations. This technique offers the researcher two advantages. First, by sampling behavior in several different situations, you are able to determine whether the behavior in question changes as a function of the context in which you observed it. For example, a researcher might use situation sampling to determine whether the amount of personal space people prefer differs from one culture to another or from one geographic area of a country to another.

The second advantage of the situation-sampling technique involves the fact that researchers are likely to observe different participants in the different situations. Because different individuals are observed, our ability to generalize any behavioral consistencies we notice across the various situations is increased. Because Gibson et al. (2000) made observations at four different ATMs and obtained the same results at each one, they could not attribute their findings to the customers at one specific machine.

Even if you have decided to time sample or situation sample, there is still another major decision you must make before you actually conduct your research project. You need to decide whether to present the results in a *qualitative* or *quantitative* manner. If you choose the qualitative approach, your report will consist of a description of the behavior in question (a *narrative record*) and the conclusions prompted by this description. Such narrative records can be in the form of written or tape-recorded notes that you make during or immediately after observing the behavior. Video recordings also are frequently made. If you write or tape record notes after the behavior has occurred, you should write or record them as soon as possible. In all narrative records, the language and terms used should be as clear and precise as possible, and the observer should avoid making speculative comments.

If your research plans call for a quantitative or numerical approach, you will need to know how you are going to measure the behavior under investigation and how you will

Situation sampling
Observing the same behavior in different situations.

analyze these measurements. We will have more to say about measurement and analysis in Chapter 9.

Using More Than One Observer: Interobserver Reliability Another consideration we must deal with in the case of observational research is whether we will use one or more observers. As the good detective knows, there are two main reasons for using more than one observer: First, one observer may miss or overlook a bit of behavior, and second, there may be some disagreement concerning exactly what was seen and how it should be rated or categorized. More than one observer may be needed even when videotape is used to preserve the complete behavioral sequence; someone has to watch the videotape and rate or categorize the behavior contained there.

When two individuals observe the same behavior, it is possible to see how well their observations agree. The extent to which the observers agree is called **interobserver reliability**. Low interobserver reliability indicates that the observers disagree about the behavior(s) they observed; high interobserver reliability indicates agreement. Such factors as fatigue, boredom, emotional and physical state, and experience can influence interobserver reliability. If both observers are well rested, interested in their task, and in good physical and emotional health, high interobserver reliability should be obtained. An observer's physical, emotional, and attitudinal state can be monitored and dealt with easily. Additionally, the need for training observers should be considered. The importance of thorough training, especially when observing complex and subtle behaviors, cannot be stressed too much. Such training should include clear, precise definitions of the behavior(s) to be observed. The trainer should provide concrete examples of positive and negative instances of the behavior in question—if at all possible.

Even if you follow all of these guidelines, it may be difficult to obtain high interobserver reliability. In some instances the problem may reside in the type of behavior being observed and how the observers code the behaviors. For example, think of how difficult it is for two observers to agree on (a) instances of "empathy," "shame," and "self-consciousness" in 2-year-old children, (b) the difference between aggression and teasing in 10-year-olds, and (c) whether a 6-week-old infant has smiled. These examples point to the difficulty in obtaining, and the importance of having, interobserver reliability.

How can you measure interobserver reliability? A simple technique involves determining the number of agreements between the two observers and the number of opportunities the observers had to agree. Once these numbers have been determined, they are used in the following formula:

$$\frac{\text{number of times observers agree}}{\text{number of opportunities to agree}} \times 100 = \text{percentage of agreement}$$

The final calculation indicates the percentage of agreement.

Another method for obtaining interobserver reliability is to calculate the correlation (see Chapter 9) between the raters' judgments and then square the correlation and multiply by 100. The resulting figure tells us the percentage of variation that is due to observer agreement; the higher the percentage, the greater the agreement.

What is a good measure of interobserver reliability? Although there are no rules to follow, a review of articles published in the two most recent issues of several journals indicated that

Interobserver reliability

The extent to which observers agree.

all articles reporting interobserver reliability had at least 85% agreement. This figure is a guideline that indicates what journal editors and reviewers consider an acceptable minimum level for interobserver reliability.

Correlational Studies

Correlational study

Determination of the relation between two variables.

In its basic form a **correlational study** involves the measurement and determination of the relation between two variables (hence the term *co-relation*). In terms of control, empirical measurement, and statistical analysis, a correlational study is likely to be more rigorous than one of the descriptive methods we've just considered. In order to understand the intent and purpose of a correlational study, we need to review some basic facts about correlations.

The Nature of Correlations

Positive correlation As scores on one variable increase, scores on the second variable also increase.

One of three basic patterns may emerge when a correlation is calculated. The two variables may be **positively correlated**: As one variable *increases*, scores on the other variable also *increase*. For example, test scores are positively correlated if a student who makes a low score on Test 1 also scores low on Test 2, whereas a student who scores high on Test 1 also scores high on Test 2. Likewise, height and weight are positively related; in general, the taller a person is, the more he or she weighs.

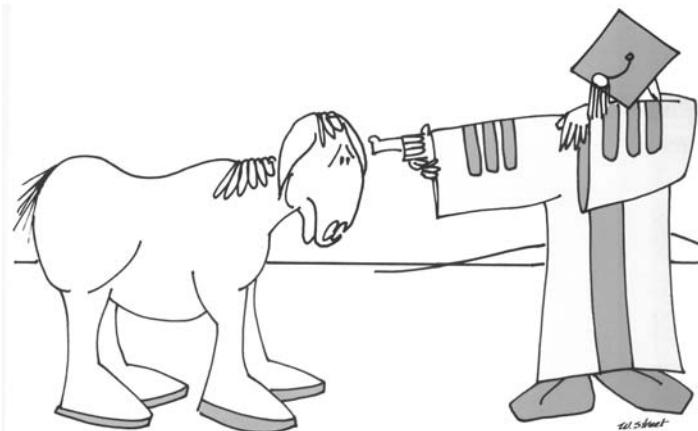
Negative correlation As scores on one variable increase, scores on the second variable decrease.

Two variables may also be negatively related. A **negative correlation** indicates that an *increase* in one variable is accompanied by a *decrease* in the second variable. For example, drinking water on a hot day and thirst are negatively related; the more water consumed, the less intense the thirst. Likewise, increasing self-esteem scores might be accompanied by decreasing anxiety scores. Thus, an individual who scores high on a self-esteem scale would have a low score on an anxiety scale, whereas an individual who scores low on the self-esteem scale would score high on the anxiety scale.

Researchers use correlations to make predictions. For example, you probably took an entrance examination, such as the American College Test (ACT) or the Scholastic Aptitude Test (SAT), when you applied for admission to college. Previous research has shown that scores on such entrance examinations are *positively correlated* with first-semester grades in college. Thus, a college or university admissions committee might use your entrance-exam score to predict how you would perform in your college classes. Obviously, the closer a correlation comes to being perfect, the better our predictions will be.



We have seen that correlations can be positive or negative in value. What has happened when we obtain a zero correlation?



(Courtesy of Warren Street)

"There's a significant NEGATIVE correlation between the number of mules and the number of academics in a state, but remember, correlation is not causation."
(Courtesy of Warren Street.)

Zero correlations indicate that the two variables under consideration are not related. High scores on one variable may be paired with low, high, or intermediate scores on the second variable, and vice versa. In other words, knowing the score on Variable 1 does not help us predict the score on Variable 2. A zero correlation may not be exactly 0. For example, a correlation of 0.03 would be considered a zero correlation by most researchers. It is important to remember that a correlation tells us only the extent to which two variables are related. Because these two variables are not under our direct control, we are still unable to make cause-and-effect statements. In short, *correlations do not imply causation*. It is possible that a third variable is involved. A rather farfetched example illustrates this point quite well. In an introductory statistics class, a student told one of your authors about a correlational study that reported the relation between the number of telephone poles erected in Australia each year for the 10 years following World War II and the yearly birthrate in the United States during the same time period. The result was a very high, positive correlation. The point of the example was to illustrate that correlation does not imply causation. It also illustrates the likely presence of a third variable. It is arguable that the increasingly better worldwide economic conditions that followed World War II encouraged both industrial development (the increase in telephone poles in Australia) and the desire to have a family (the higher birthrate in the United States).

Zero correlation Two variables under consideration are not related.

Correlational Research

We agree that it is confusing when a type of mathematical analysis and a research technique have the same name. Because the common goal is to determine the relation between two variables or factors, more than coincidence is at work. Also, you should keep in mind that a correlational research project uses the correlational mathematical technique to analyze the data gathered by the researcher. Now that we have reviewed the nature and types of correlations, let's examine an example of correlational research. Kristen Robinson, a student at John Carroll University (University Heights, OH) was interested in determining what factors were good "predictors of health in order to gain knowledge of what steps can be taken to insure [sic] health during the college years" (Robinson, 2005, p. 3). To accomplish her objective, she

administered several surveys to 60 college students (30 women and 30 men). She found that locus of control and overall college adjustment were significantly correlated with illness severity. More specifically, her data showed “that students with an internal locus of control and more complete college adjustment had less severe episodes of illness than did those with an external locus of control and incomplete college adjustment” (Robinson, 2005, p. 6).

Researchers use correlational studies when data on two variables are available but they can only measure, rather than manipulate, either variable. For example, Jessica Serrano-Rodriguez, Sara Brunoli, and Lisa Echols (2007), students at the University of San Diego, examined the relation between religious attitudes and approval of organ donation. Likewise, Adam Torres, Christy Zenner, Daina Benson, Sarah Harris, and Tim Koberlein (2007), students at Boise State University (Boise, ID), calculated the correlations between self-esteem and factors such as perceived academic abilities, family supportiveness, writing skills, and so forth, and college attendance. Although correlational investigations such as these can determine the degree of relation that exists between two variables, they are not able to offer a cause-and-effect statement concerning these variables.

■ REVIEW SUMMARY

1. **Descriptive research methods** are nonexperimental procedures for acquiring information. They do not involve the manipulation of variables.
2. Some researchers make use of archival and previously recorded sources of data. Although the use of such data avoids biasing participants' responses, this approach suffers from lack of generalizability, selective deposit, and selective survival.
3. Although observational techniques do not involve the direct manipulation of variables, they do allow the researcher to observe the behavior(s) of interest directly.
4. **Naturalistic observation** involves directly observing behaviors in the natural environment. In these studies the observer should remain unobtrusive in order to avoid a **reactance** or **reactivity effect** on the part of the participants.
5. Because the observer may not be able to observe all behaviors at all times, decisions concerning which behaviors to observe, as well as when and where to observe them, must be made.
6. **Time sampling** involves making observations at different time periods, whereas **situation sampling** involves observing the same behavior(s) in several different situations.
7. The use of more than one observer may be desirable to avoid missing important observations and to help resolve disagreements concerning what was or was not observed. **Interobserver reliability** refers to the degree to which the observers agree.
8. A **correlational study** involves the measurement and determination of the relation between two variables.
9. Two variables are **positively correlated** when an increase in one variable is accompanied by an increase in the other variable.
10. Two variables are **negatively correlated** when an increase in one variable is accompanied by a decrease in the other variable.
11. A **zero correlation** exists when a change in one variable is unrelated to changes in the second variable.

Check Your Progress

1. What is the reactance effect? How is it avoided by the use of archival sources of data?
2. Selective deposit is a problem associated with
 - a. case studies
 - b. naturalistic observation
 - c. cause-and-effect research
 - d. archival research
3. You want to write a history of the psychology department at your school, so you look through old catalogs, departmental correspondence, and musty file folders you find in the basement. You are conducting
 - a. archival research
 - b. a case study
 - c. an experiment
 - d. a participant observation study
4. The extent to which experimenters agree about a particular bit of data best describes
 - a. archival symmetry
 - b. interobserver reliability
 - c. concordance of naturalistic observation
 - d. participant–observer ratio
5. Why are time sampling and situation sampling employed?
6. What is interobserver reliability? How is it calculated?
7. Which of the following would probably be *negatively* correlated?
 - a. number of days present in class and exam grades
 - b. height and weight of college students
 - c. number of difficult classes taken and semester GPA
 - d. overall college GPA and graduate school GPA

Ex Post Facto Studies

Can we conduct experimental research on variables that we cannot control or manipulate? Yes, we can conduct research on such variables, but we must be cautious in drawing conclusions. When we work with independent variables (IVs) that we cannot or do not manipulate, we are conducting an **ex post facto study**. Ex post facto is a Latin phrase meaning “after the fact.” When we conduct an ex post facto study, we are using an IV “after the fact”—it has already varied before we arrived on the scene. A great deal of detective work would seem to fall into this category. Because the experimenter has no control over administering the IV, let alone determining who receives this variable and under what conditions it is administered, the ex post facto study clearly qualifies as a descriptive research technique. However, it does have some properties in common with experimental methods.

Let’s look at an example of student research using an ex post facto approach. Carolyn Licht of Marymount Manhattan College in New York City evaluated occupational stress as a function of the sex of the participants and whether the participants worked for a nonprofit or for-profit organization. Her results showed “that employees perceive more occupational stress in nonprofit than in for-profit organizations” and that “men report more stress than women in most situations” (Licht, 2000, p. 46).

Ex post facto study A study in which the variable(s) to be studied are selected after they have occurred.



What causes Licht's research to fall into the ex post facto category?

Because Licht had no control over the sex of the participants or the type of organization the participants worked for, these variables put this project in the ex post facto category. In another ex post facto study, Christy Zenner (a student at Boise State University) and her faculty supervisor, Mary Pritchard, were interested in what college students know about breast cancer and eating disorders. They hypothesized that female students would have greater knowledge of these topics than male students. Their data supported these predictions (Zenner & Pritchard, 2007). In the next section, we will consider research using surveys, questionnaires, tests, and inventories; such research is quite popular.

Surveys, Questionnaires, Tests, and Inventories

Whether they are used in experimental or nonexperimental research, investigators frequently use surveys, questionnaires, tests, and inventories to assess attitudes, thoughts, and emotions or feelings. One reason surveys and questionnaires are popular is that they appear to be quite simple to conduct; when you want to know how a group of individuals feel about a particular issue, all you have to do is ask them or give them a test. As we will see, appearances can be quite deceptive; it is not as easy to use this technique as it appears. What type of instrument is best suited for our particular project? There are numerous choices available to us. We first consider surveys and questionnaires, then turn our attention to tests and inventories.

Surveys and Questionnaires

Surveys typically request our opinion on some topic or issue that is of interest to the researcher. There are two basic types of surveys: descriptive and analytic. Although we will discuss these types separately, you should keep in mind that some surveys can serve both purposes.

Types of Surveys The purpose of your research project will determine the type of survey you choose to administer. If you seek to determine what percentage of the population has a certain characteristic, holds a certain opinion, or engages in a particular behavior, then you will use a **descriptive survey**. The Gallup Polls that evaluate voter preferences and the Nielsen television ratings are examples of descriptive surveys. When a researcher uses this type of survey, there is no attempt to determine what the relevant variables are and how they may relate to the behavior in question. The end product is the description of a particular characteristic or behavior of a sample with the hope that this finding is representative of the population from which the sample was drawn.

Descriptive survey

Seeks to determine the percentage of the population that has a certain characteristic, holds a particular opinion, or engages in a particular behavior.



© 2006 Harley Schwadron from cartoonbank.com.
All rights reserved.

The **analytic survey** seeks to determine what the relevant variables are and how they might be related. For example, Amanda Gray, a student at Agnes Scott College in Decatur, Georgia, and her faculty supervisor, Jennifer Lucas, used the Subjective Impediment Scale (Novaco, Stokols, & Milanesi, 1990) to determine “automobile commuters’ subjective perceptions of travel impedance and their driver stress” (Gray & Lucas, 2001, p. 79). They found that one of the key determinants of commuters’ stress level is the *perception* of impedance, regardless of whether it actually occurred.

What about conducting research in areas in which there are no readily available surveys? To address the issues you are interested in, you may have to construct your own survey. We *strongly* encourage you to exhaust every possibility before you attempt to construct your own survey. They look like they are easy to construct, but nothing could be further from the truth.

If you have decided that the only option is to construct a survey, you will have to choose the questions for your analytic survey *very carefully* before the survey is administered. In fact, it will probably be necessary to do some pilot testing of the analytic survey before you use it in a full-scale investigation. **Pilot testing** refers to the testing and evaluating that is done in advance of the complete research project. During this preliminary stage the researcher tests a small number of participants and may even use in-depth interviews to help determine the type of questions that should appear on the final survey instrument.

Developing a Good Survey or Questionnaire A good survey or questionnaire, one that measures the attitudes and opinions of interest in an unbiased manner, is not developed overnight; considerable time and effort typically go into its construction. When you have decided exactly what information your research project seeks to ascertain, you should follow several steps in order to develop a good survey or questionnaire. These steps appear in Table 4-1.

Analytic survey Seeks to determine the relevant variables and how they are related.

Pilot testing Preliminary, exploratory testing that is done prior to the complete research project.

TABLE 4-1 Steps in Developing a Good Survey or Questionnaire

- | | |
|---------------|--|
| Step 1 | Decide what type of instrument to use. How will the information be gathered? |
| Step 2 | Identify the types of questions to use. |
| Step 3 | Write the items: They should be clear, short, and specific. |
| Step 4 | Pilot test and seek opinions from knowledgeable others. |
| Step 5 | Determine the relevant demographic data to be collected. |
| Step 6 | Determine administration procedures and develop instructions. |

The first step is to determine how you will obtain the information you seek. Will you use a mail survey? Will your project involve the use of a questionnaire that is administered during a regular class session at your college or university? Will trained interviewers administer your questionnaire in person, or will a telephone interview be conducted? Decisions such as these will have a major impact on the type of survey or questionnaire you will develop.

After you have decided which type of instrument you will develop, in the second step you can give attention to the *nature* of the questions that you will use and the type of responses the participants will make to these questions. Among the types of questions that are frequently used in surveys and questionnaires are the following:

1. Yes-No Questions. The respondent answers yes or no to the items.

EXAMPLE:

The thought of death seldom enters my mind.

(Source: Templer's Death Anxiety Scale; Templer, 1970)

2. Forced Alternative Questions. The respondent must select between two alternative responses.

EXAMPLE:

A. There are institutions in our society that have considerable control over me.

B. Little in this world controls me. I usually do what I decide to do.

(Source: Reid-Ware Three-Factor Locus of Control Scale; Reid & Ware, 1973)

3. Multiple-Choice Questions. The respondent must select the most suitable response from among several alternatives.

EXAMPLE:

Compared to the average student,

A. I give much more effort.

B. I give an average amount of effort.

C. I give less effort.

(Source: Modified Jenkins Activity Scale; Krantz, Glass, & Snyder, 1974)

4. Likert-Type Scales. The individual answers a question by selecting a response alternative from a designated scale. A typical scale might be the following: (5) strongly agree, (4) agree, (3) undecided, (2) disagree, or (1) strongly disagree.

EXAMPLE:

I enjoy social gatherings just to be with people.



(Source: Texas Social Behavior Inventory; Helmreich & Stapp, 1974)

- 5. Open-Ended Questions.** A question is asked to which the respondent must construct his or her own answer.

EXAMPLE:

How would you summarize your chief problems in your own words?

(Source: Mooney Problem Check List; Mooney, 1950)

Clearly, the questions you choose to use on your survey or questionnaire will directly influence the type of data you will gather and be able to analyze when your project is completed. If you choose the yes–no format, then you will be able to calculate the frequency or percentage of such responses for each question. The use of a Likert-type scale allows you to calculate an average or mean response to each question. Should you choose to use open-ended questions, you will have to either decide how to code or quantify the responses or establish a procedure for preparing a summary description of each participant’s answers.

The third step is to write the items for your survey or questionnaire. As a general rule, these questions should be clear, short, and specific; use familiar vocabulary; and be at the reading level of the individuals you intend to test. In preparing your items, you should avoid questions that might constrain the respondents’ answers. For example, you might ask your participants to rate the effectiveness of the president of the United States in dealing with “crises.” Assuming that the president has dealt with several crises, it may not be clear whether the question is referring to one particular type of crisis or another. Hence, the respondents will have to interpret the item as best they can; their interpretation may not coincide with your intended meaning. Also, researchers should avoid using questions that might bias or prejudice the respondents’ answers. A negatively worded question may result in a preponderance of negative answers, whereas a positively worded question may result in a preponderance of positive answers.



Consider this yes–no question: “Do you agree that wealthy professional athletes are overpaid?” What is wrong with this question, and how can it be improved?

By indicating that professional athletes are wealthy, you have created a mental set that suggests that they may be overpaid. Thus, your respondents may be biased to answer yes. Also, using the word *agree* in the question may encourage yes answers. If the researcher rewrites the question as follows, it is less likely to bias the respondents’ answers: “Do you believe professional athletes are overpaid?”

The fourth step is to pilot test your survey or questionnaire. It is important to ask others, especially professionals who have expertise in your area of research interest, to review your items. They may be able to detect biases and unintended wordings that you had not considered. It will also be helpful at this preliminary stage to administer your questionnaire to several individuals and then discuss the questions with them. Often there is nothing comparable to the insights of a participant who has actually completed a testing instrument. Such insights can be invaluable as you revise your questions. In fact, you may find it necessary to pretest and revise your survey in this manner several times before developing a final draft.

The fifth step involves a consideration of the other relevant information that you want your participants to provide. Frequently, such information falls under the heading of **demographic data**, which may include such items as age, sex, annual income, size of community, academic major, and academic classification. For example, in Chapter 5 we will examine research on sex differences in spatial task performance conducted by Robyn Scali (Scali & Brownlow, 2001). In their report Scali and Brownlow indicated, "Because training and prior experience may have an effect on spatial ability, we assessed experience with art, math, sports, and similar tasks" (Scali & Brownlow, 2001, p. 6); clearly, they were attentive to demographic needs.

Although the need for this step may seem obvious, it is important to review these items carefully to ensure that you have not forgotten to request a vital bit of information. We cannot begin to tell you how many survey projects designed to evaluate male–female differences were less than successful because they failed to include an item that requested the participant's sex!

The final step is to specify clearly the procedures that will be followed when the survey or questionnaire is administered. If the survey is self-administering, what constitutes the printed instructions? Are they clear, concise, and easy to follow? Who will distribute and collect the informed consent forms and deal with any questions that may arise? If your survey or questionnaire is not self-administering, then you must prepare an instruction script. The wording of this set of instructions must be clear and easily understood. Whether you present these instructions in a face-to-face interview, over the telephone, or in front of a large class of students, you (or the person giving them) must thoroughly practice and rehearse them. The researcher in charge of the project must be sure that all interviewers present the instructions in the same manner on all occasions. Likewise, questions raised by the participants must be dealt with in a consistent manner.

As we saw, the final step in creating a good survey involves a determination of the administration procedures. Because these choices are crucial to the success of this type of research, we will examine the three basic options—mail surveys, personal interviews, and telephone interviews—in more detail.

Mail Surveys Most likely you have been asked to complete a survey you received in the mail. This popular technique is used to gather data on issues that range from our opinions on environmental problems to the type of food we purchase.

One advantage of sending surveys through the mail is that the researcher does not have to be present while the survey is being completed. Thus, surveys can be sent to a much larger number of participants than a single researcher could ever hope to contact in person.

Although it is possible to put a survey in the hands of a large number of respondents, there are several disadvantages associated with this research strategy. First, the researcher cannot be sure who actually completes the survey. Perhaps the intended respondent was too busy to

Demographic data

Information about participants' characteristics such as age, sex, income, and academic major.

complete the survey and asked a family member or friend to finish it. Hence, the time and trouble spent in creating a random sample from the population of interest may be wasted.

Even if the intended respondent completes the survey, there is no guarantee that the respondent will answer the questions in the same order in which they appeared on the survey. If the order of answering questions is relevant to the project, then this drawback may be a major obstacle to the use of mail surveys.

The low return rate associated with the use of mail surveys highlights another problem. In addition to disappointment and frustration, low return rates suggest a potential bias in the researcher's sample. What types of individuals returned the surveys? How did they differ from those individuals who did not return the surveys? Were they the least (most) busy? Were they the least (most) opinionated? We really do not know, and as the response rate drops lower, the possibility of having a biased sample increases. What constitutes a good response rate to a mail survey? It is not uncommon to have a response rate of from 25% to 30% to a mail survey; response rates of 50% and higher are considered quite acceptable.



Assume that you are planning to conduct a mail survey project. You are concerned about the possibility of having a low response rate and want to do everything to ensure the return of your surveys. What can you do to increase your response rate?

Researchers have developed strategies such as these to increase the response rates of mail surveys:

1. The initial mailing should include a letter that clearly summarizes the nature and importance of the research project, how the respondents were selected, and the fact that all responses are confidential. You should include a prepaid envelope for the return of the completed survey.
2. It may be necessary to send an additional mailing to your respondents. Because the original survey may have been misplaced or lost, it is important to include a replacement. One extra mailing may not be sufficient; you may find it necessary to send two or three requests before you achieve an acceptable response rate. These extra mailings are typically sent at 2- to 3-week intervals.

Not all researchers endorse the use of mail surveys. Low response rates, incomplete surveys, and unclear answers are among the reasons that cause some researchers to use direct interviews to obtain data. These interviews may be done in person or over the telephone.

Personal Interviews When a trained interviewer administers a survey in a respondent's home, the response rate climbs dramatically. It is not uncommon to have a 90% completion rate under these circumstances. In addition to simply increasing the response rate, the trained interviewer is able to cut down on the number of unusable surveys by clarifying ambiguous questions, making sure that all questions are answered in the proper sequence, and generally assisting with any problems experienced by the respondents.

Although this technique offers some advantages when compared to the mail survey, there are drawbacks. First, the potential for considerable expenditure of time and money exists.

Time has to be devoted to the training of the interviewers. Once trained, they will have to be paid for their time on the job. Second, the fact that an individual is administering the survey introduces the possibility of interviewer bias. Some interviewers may present some questions in a more positive (or negative) manner than do other interviewers. Only careful and extensive training of all interviewers to present all items in a consistent, neutral manner can overcome this potential difficulty. Finally, the prospect of administering surveys in the home is becoming less appealing and feasible. In many instances no one is at home during the day, and an increasing number of people are not willing to sacrifice their leisure evening hours to complete a survey. Additionally, the high crime rates in many urban areas discourage face-to-face interviewing; in its place, many investigators have turned to telephone interviewing.

Telephone Interviews In addition to overcoming several of the problems associated with personal interviews and mail surveys, telephone interviewing offers several advantages. For example, the development of random-digit dialing allows researchers to establish a random sample with ease: The desired number of calls is specified, and the computer does the rest. It is noteworthy that a random sample generated in this manner will contain both listed and unlisted telephone numbers because the digits in each number are selected randomly. With over 95% of all households in the United States currently having telephones, previous concerns about creating a biased sample consisting of only households having telephones seem largely unfounded.

Computer technology also has increased the desirability of conducting telephone interviews. For example, it is now possible to enter the participant's responses directly as they are being made. Hence, the data are stored directly in the computer and are ready for analysis at any time.

Despite these apparent advantages, telephone interviews do have potential drawbacks. Even though technology has assisted telephone researchers, it also has provided an obstacle. Many households are now equipped with answering machines or caller ID that allow incoming calls to be screened or blocked. Even if the call is answered, it is easier to say no to an unseen interviewer on the telephone than to a person at your front door. Additionally, people who have added call blocking to their phone service will not be included in the sample. These three situations lower the response rate and raise the possibility of a biased sample.

The use of the telephone also prohibits the use of visual aids that might serve to clarify certain questions. In addition, because the telephone interviewer cannot see the respondent, it is not possible to evaluate nonverbal cues such as facial expressions, gestures, and posture. Such cues might suggest that a certain question was not completely understood or that an answer is in need of clarification.

Not being in face-to-face contact with the respondent also makes it more difficult to establish rapport. Hence, telephone respondents may not be as willing to participate in the survey. This potential lack of willingness has led to the use of shorter survey instruments.

Although surveys and questionnaires are popular research tools with many investigators, there are other ways to gather data. Since the late 1800s when Sir Francis Galton (1822–1911) attempted to evaluate people's ability or intelligence by measuring physical attributes such as reaction time or visual acuity, psychologists have developed a large number of tests and inventories for a wide variety of purposes.

Tests and Inventories

Unlike surveys and questionnaires, which evaluate opinions on some topic or issue, tests and inventories are designed to assess a specific attribute, ability, or characteristic possessed by

the individual being tested. In this section we look at the characteristics of good tests and inventories and then discuss three general types of tests and inventories: achievement, aptitude, and personality.

Characteristics of Good Tests and Inventories Unlike surveys and questionnaires, the researcher is less likely to be directly involved with the development of a test or inventory. Because their development and pilot testing has already taken place, you will need to scrutinize the reports concerning the development of each test or inventory you are considering. A good test or inventory should possess two characteristics: It should be valid, and it should be reliable.

Validity A test or inventory has **validity** when it actually measures what it is supposed to measure. If your research calls for a test that measures spelling achievement, you want the instrument you select to measure that ability, not another accomplishment, such as mathematical proficiency.

There are several ways to establish the validity of a test or inventory.

Content validity indicates that the test items actually represent the type of material they are supposed to test. Researchers often use a panel of expert judges to assess the content validity of test items. Although the more subjective evaluation of such judges may not lend itself to a great deal of quantification, their degree of agreement, known as **interrater reliability**, can be calculated. Interrater reliability is similar to interobserver reliability. The main difference is that interrater reliability measures agreement between judgments concerning a test item, whereas interobserver reliability measures agreement between observations of behavior.

We can establish **concurrent validity** when we already have another measure of the desired trait or outcome and can compare the score on the test or inventory under consideration with this other measure. For example, the scores made by a group of patients on a test designed to measure aggression might be compared with a diagnosis of their aggressive tendencies made by a clinical psychologist. If the test and the clinical psychologist rate the aggressiveness of the patients in a similar manner, then concurrent validity for the test has been established.

Often our second measure may not be immediately accessible. When the test score is to be compared with an outcome that will occur in the future, the researcher is attempting to establish the **criterion validity** of the test. Thus, criterion validity refers to the ability of the test or inventory to predict the outcome or criterion. For example, it is the desired outcome that college entrance examinations such as the SAT and ACT predict first-semester performance in college. To the extent that these tests are successful at predicting first-semester GPAs, their criterion validity has been established.

Reliability Once we have determined that a particular test is valid, we also want to make sure that it is reliable. **Reliability** refers to the extent to which the test or inventory is consistent in its evaluation of the same individuals over repeated administrations. For example, if we have

Validity The extent to which a test or inventory measures what it is supposed to measure.

Content validity The extent to which test items actually represent the type of material they are supposed to represent.

Interrater reliability Degree of agreement among judges concerning the content validity of test or inventory items.

Concurrent validity Degree to which the score on a test or inventory corresponds with another measure of the designated trait.

Criterion validity Established by comparing the score on a test or inventory with a future score on another test or inventory.

Reliability Extent to which a test or inventory is consistent in its evaluation of the same individuals.

developed a test to measure aptitude for social work, we would want individuals who score high (or low) on our test on its first administration to make essentially the same score when they take the test again. The greater the similarity between scores produced by the same individuals on repeated administrations, the greater the reliability of the test or inventory.

Test-retest procedure

Determination of reliability by repeatedly administering a test to the same participants.

Reliability is typically assessed through the test-retest or split-half procedures. When the **test-retest procedure** is used, the test is simply given a second time and the scores from the two tests are compared; the greater the similarity, the higher the reliability.



On the surface, the test-retest procedure appears to be quite straightforward and reasonable; however, there may be a problem with establishing reliability. What is the problem?

The main problem with the test-retest procedure comes from the fact that the participants are repeatedly administered the same test or inventory. Having already taken the test or inventory, the individuals may remember the questions and answers the next time the instrument is administered. Their answers may be biased by the previous administration. If a lengthy time period elapses between administrations, the participants might forget the questions and answers, and this familiarity problem might be overcome. On the other hand, a lengthy time period may influence reliability in yet another manner. An extended time period allows the participants to have numerous experiences and learning opportunities between administrations. These experiences, called *history effects* (see Chapter 8), may influence their scores when the participants take the test or inventory again. Hence, the reliability measure may be influenced by the experiences that intervene between the two testing sessions.

Split-half technique

Determination of reliability by dividing the test or inventory into two subtests and then comparing the scores made on the two halves.

It is possible to overcome the problems of test familiarity and lengthy time periods separating administrations by using the split-half approach. The **split-half technique** of establishing reliability involves dividing a test or inventory into two halves or subtests and then administering them to the same individuals on different occasions or by administering the entire test and then splitting it into two halves. Because the questions that comprise the two subtests came from the same test, it is assumed that they are highly related to each other if the test is reliable. Typically, the questions that comprise these two subtests are selected randomly or in some pre-determined manner, such as odd-even. The higher the degree of correspondence between scores on the two subtests, the greater the reliability of the overall test from which they were selected.

Having determined that a test or inventory should be valid and reliable, we now examine several types of these instruments that are currently used for research and predictive purposes.

Achievement test

Designed to evaluate an individual's level of mastery or competence.

Types of Tests and Inventories **Achievement tests** are given when an evaluation of an individual's level of mastery or competence is desired. For example, doctors must pass a series of medical board examinations before they are allowed to practice medicine, and lawyers must pass the bar examination before they are allowed to practice law. The score that distinguishes passing from failing

determines the minimum level of achievement that must be attained. You can probably think of many achievement tests you have taken during your life.

At many colleges and universities the counseling center or career development office offers students the opportunity to take an aptitude test to assist them in selecting a major or making a career choice. An **aptitude test** is used to assess an individual's ability or skill in a particular situation or job. For example, the Purdue Pegboard Test is often administered to determine aptitude for jobs that require manual dexterity. According to Anastasi (1988), "[T]his test provides a measure of two types of activity, one requiring gross movements of hands, fingers, and arms, and the other involving tip-of-the-finger dexterity needed in small assembly work" (p. 461). Similarly, if you are planning to attend graduate school, you will probably be required to take the Graduate Record Examination (GRE). For most graduate schools, the two most important scores on the GRE are the verbal and quantitative subtests. These scores represent measures of your aptitude to complete successfully verbal and quantitative courses on the graduate level.

Aptitude test Designed to assess an individual's potential ability or skill in a particular job.

The **personality test or inventory** measures a specific aspect of an individual's motivational state, interpersonal capability, or personality (Anastasi & Urbina, 1997). The use of a personality inventory in research is exemplified by a project reported by Dana Bodner and C. D. Cochran, students at Stetson University in DeLand, Florida, and their faculty advisor, Toni Blum (Bodner, Cochran, & Blum, 2000). The purpose of this study was to validate a scale, the General Unique Invulnerability Scale (GUI), which measures general optimism about unique invulnerability (the belief that you will not experience negative events or misfortunes). These researchers administered the GUI to a sample of 40 skydivers and a sample of 40 college students and found that skydivers had higher GUI scores than did the college students. It is not surprising that skydivers would have a stronger belief in unique invulnerability than the typical college student who does not engage in this risky behavior.

Personality test or inventory Measures a specific aspect of the individual's motivational state, interpersonal capability, or personality.

Sampling Considerations and Basic Research Strategies

Having completed our review of surveys, questionnaires, tests, and inventories, we conclude this chapter with a consideration of two more issues that all researchers must consider when they conduct a project: sampling and basic research strategies. Sampling deals with the question of who will participate in your research project and whether the participants are a representative group. Once you have dealt with sampling issues, then you must decide on your main research strategy. The main strategies used by researchers are the single-strata, cross-sectional, and longitudinal approaches.

Sampling

Assume that you want to determine which of two new titles for the college newspaper—*The Wilderbeast* (named after your school mascot) or *The Observer*—appeals most to the student body. You ask the 36 students in your senior-level biopsychology course and find that 24 prefer *The Observer* and 12 prefer *The Wilderbeast*. You report your findings to the publications advisory board and recommend that *The Observer* be chosen as the new title.



Should the publications board accept your recommendation? Are there some reasons to question your findings?

The publications advisory board should not accept your recommendation. Unlike Sherlock Holmes, who said, “I presume nothing” (Doyle, 1927, p. 745), you seem to have made some assumptions. The main problem with your data concerns the students you surveyed. Is your senior-level biopsychology class representative of the entire campus? The answer to this question must be a resounding no; only senior-level psychology majors take this class. Moreover, a quick check of the class roster indicates that the majority (67%) of the students in your class are women. Clearly, you should have selected a group of students more representative of the general student body at your college.

Population The complete set of individuals or events.

Sample A group that is selected to represent the population.

Random sample A sample in which every member of the population has an equal likelihood of being included.

Random sampling without replacement
Once chosen, a score, event, or participant cannot be returned to the population to be selected again.

Random sampling with replacement Once chosen, a score, event, or participant can be returned to the population to be selected again.

We shall designate the general student body as the **population** or the complete set of individuals or events we want to represent. The group that we select to represent the population is called the **sample**. When every member of the population has an equal likelihood of being selected for inclusion in the sample, we have created a **random sample**.

How would you obtain a random sample of students on your campus to take the newspaper title survey? Computer technology has made this task quite simple; you simply indicate the size of the sample you desire, and the computer can be programmed to randomly select a sample of that size from the names of all currently enrolled students. Because a name is not eligible to be chosen again after it has been selected, this technique is called **random sampling without replacement**. If the chosen item can be returned to the population and is eligible to be selected again, the procedure is termed **random sampling with replacement**. Because psychologists do not want the same participant to appear more than once in a group, random sampling without replacement is the preferred technique for creating a random research sample.

Suppose a sample of 80 students has been randomly selected from the entire student body by the computer and you are examining the printout. Even though you selected this sample randomly, it also has some apparent problems. Just by chance the majority of the students selected are freshmen and sophomores. Moreover, the majority of students in this sample are men. The views on the two newspaper titles held by this group of randomly selected students may not be much more representative than our original sample. What can we do to produce an even more representative sample?

There are two techniques that we can use to increase the representativeness of our sample. The first procedure is quite simple: We can select a larger sample. Generally speaking, the larger the sample, the more representative it will be of the population. If we randomly selected 240 students, this larger sample would be more representative of the general student body than our original sample of 80 students.

Although larger samples may be more like the population from which they are drawn, there is a potential drawback. Larger samples mean that more participants will need to be tested. In our project dealing with the two newspaper titles, the testing of additional participants may not present any major problems. However, if we were administering a lengthy questionnaire or paying participants for their participation, increasing the number of participants might create unmanageable time or financial obstacles.

If simply increasing the sample size does not offer a good solution to the problem of achieving a representative sample, the researcher may want to use stratified random sampling. **Stratified random sampling** involves dividing the population into subpopulations or strata and then drawing a random sample from one or more of these strata. For example, one logical subdivision of a college student body would be by classes: freshmen, sophomores, juniors, and seniors. You could then draw a random sample from each class. How many students will be included in each stratum? One option would be for each stratum to contain an equal number of participants. Thus, in our newspaper title project we might include 20 participants from each academic classification. A second option would be to sample each stratum in proportion to its representation in the population. If freshmen comprise 30% of the student body, then our random sample would contain 30% first-year students. What about the number of men and women sampled in each stratum? We could have equal numbers, or we could sample in proportion to the percentage of men and women in each stratum. As you have probably surmised, the use of stratified random sampling indicates that you have considerable knowledge of the population in which you are interested. Once you have this knowledge, you can create a sample that is quite representative of the population of interest. A word of caution is in order, however. Although it may be tempting to specify a number of characteristics that your sample must possess, you can carry this process too far. If your sample becomes too highly specified, then you will be able to generalize or extend your results only to a population having those very specific characteristics.

One stratum that frequently appears in research reports is the subject or participant pool, which is used by many colleges and universities. Here, students, typically enrolled in introductory psychology, have the option (among others) of participating in psychological research in order to fulfill a course requirement. After students have volunteered to participate in a research project, the researcher can randomly assign them to specific groups or treatment conditions. For example, in the research on the wording of the informed consent document done by Burkley et al. (2000), which we considered in Chapter 2, the authors stated,

Twenty-five undergraduate university psychology students (2 men, 23 women) volunteered to participate. Participants were randomly assigned to either the control or experimental group. Participants received class credit for their involvement in the study. (p. 44)

Stratified random sampling Random samples are drawn from specific subpopulations or strata of the general population.

Basic Research Strategies

Even though you have chosen your sample, you cannot simply rush out to start testing participants. You need to give some thought to the research question and how you can best conduct your project in order to answer that question. There are three basic approaches you can adopt: single-strata, cross-sectional, and longitudinal.

Single-strata approach

Gathering data from a single stratum of the population of interest.

Cross-sectional research

Comparison of two or more groups during the same, rather limited, time period.

Longitudinal research project

Obtaining research data from the same group of participants over an extended period of time.

Cohort

A group of individuals born during the same time period.

The **single-strata approach** seeks to acquire data from a single, specified segment of the population. For example, a particular Gallup Poll may be interested only in the voting preferences of blue-collar workers. Hence, a sample composed only of individuals from this stratum would be administered a voter-preference survey. This approach typically seeks to answer a rather specific research question.

When the single-strata approach is broadened to include samples from more than one stratum, a cross-sectional approach is being employed. **Cross-sectional research** involves the comparison of two or more groups of participants during the same, rather limited, time span. For example, a researcher may want to compare voter preferences of different age groups. To acquire this information, random samples of voters ages 21, 31, 41, 51, 61, and 71 are obtained and their responses to a voter-preference survey are compared.

Perhaps the researcher wants to obtain information from a group of participants over an extended period of time. In this instance a **longitudinal research project** would be conducted. First, the researcher would obtain a random sample from the population of interest; then this sample would complete an initial survey or test. The same participants would then be contacted periodically to determine what, if any, changes had occurred during the ensuing time in the behavior of interest. This group of individuals, born in the same time period and repeatedly surveyed or tested, is called a **cohort**. For example, a researcher might be interested in changes in the degree of support for environmental conservation that may occur as individuals grow older. To evaluate such changes, a group of grade-school children is randomly selected. Every five years, all members of this cohort are contacted and administered an environmental conservation survey. Table 4-2 allows you to compare these three different research strategies.

TABLE 4-2 Single-Strata, Cross-Sectional, and Longitudinal Research Strategies

Single Strata

1st-graders
2nd-graders
3rd-graders
4th-graders
5th-graders
6th-graders

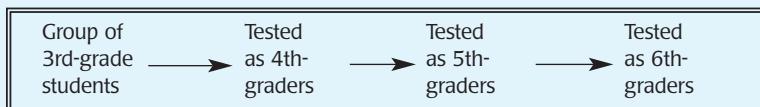
A group is selected from **one** stratum

Cross-Sectional

1st-graders
2nd-graders
3rd-graders
4th-graders
5th-graders
6th-graders

- sample
- sample
- sample
- sample
- sample
- sample

A sample is selected from each stratum.

Longitudinal

One group is tested repeatedly over a period of time.

It is important that you remember that sampling concerns and decisions about your basic research strategy apply to both nonexperimental and experimental research projects. Moreover, our concerns with sampling and research strategy do not end at this point. For example, we will have more to say about the importance of randomization when we discuss experimental control in Chapter 6. Likewise, once we have decided on our basic research strategy, we will have to consider such details as how many groups of participants must be tested to answer our research question. These topics fall under the general heading of experimental design; we begin our discussion of this topic in Chapter 10.

REVIEW SUMMARY

1. In an **ex post facto study** the variables have been experienced before they are examined by the researcher; therefore, control and manipulation of variables cannot be accomplished.
2. Surveys, questionnaires, tests, and inventories are used to assess attitudes, thoughts, and emotions or feelings.
3. **Descriptive surveys** seek to determine the percentage of a population that has a certain characteristic, holds a certain opinion, or engages in a particular behavior. **Analytic surveys** seek to determine the variables that are relevant in a situation and their relation.
4. The steps to be completed in developing a good survey or questionnaire include considering the type of instrument to be developed, determining the types of questions to be used, writing the items, **pilot testing**, determining the relevant demographic data to be collected, and deciding on administration procedures.
5. **Demographic data** include relevant information, such as gender, age, income, and educational level, about the participants.
6. Mail surveys can be sent to a large number of potential respondents; however, the researcher cannot be sure who actually completed the survey or in what order they completed the questions. Low response rates for mail surveys can be improved by stressing the importance of the project and by sending additional mailings.
7. Personal interviews yield a higher rate of completed surveys, but they are costly in terms of time and money. The increase in the number of working families and the escalating crime rate in urban areas have made the use of personal interviews less desirable.
8. Telephone interviews allow the researcher to reach a large number of respondents more efficiently than personal interviews and mail surveys. On the other hand, answering machines and the inability to see nonverbal cues are drawbacks to this approach.
9. Tests and inventories should be **valid** (measure what they are supposed to measure) and **reliable** (be consistent in their evaluation).
10. **Validity** may be established by the content, concurrent, and criterion methods.
11. The **test-retest** and **split-half procedures** are used to establish **reliability**.
12. **Achievement tests** evaluate level of mastery or competence. **Aptitude tests** assess an individual's ability or skill in a particular situation or job. A **personality test or inventory** measures a specific aspect of the individual's motivational state, interpersonal capability, or personality.

13. The **sample** of individuals who complete a survey, questionnaire, test, or inventory should be representative of the **population** from which it is drawn.
14. When a **random sample** is selected, every member of the population has an equal likelihood of being selected. When **random sampling without replacement** is used, an item cannot be returned to the population once it has been selected. When **random sampling with replacement** is used, selected items can be returned to the population to be selected again.
15. **Stratified random sampling** involves dividing the population into subpopulations or strata and then drawing random samples from these strata.
16. Basic research strategies include investigating (a) a **single stratum** of a specified population, (b) samples from more than one stratum in a **cross-sectional** project, or (c) a single group of participants over an extended time period in a **longitudinal study**.

Check Your Progress

1. Matching
 1. descriptive survey
 2. analytic survey
 3. pilot testing
 4. demographic data
 5. cohort
 6. concurrent validity
 7. criterion validity
 - A. defines participants born in the same time period
 - B. affects ability to predict the outcome
 - C. may include age, sex, and annual income
 - D. indicates the percentage having a certain characteristic
 - E. compares scores on two separate measures
 - F. tries to determine what the relevant variables are
 - G. does testing or evaluating in advance of the complete research project
2. Describe the steps involved in creating a good survey.
3. "Working with IVs that the experimenter does not manipulate" best describes
 - a. a case study
 - b. naturalistic observation
 - c. participant observation
 - d. an ex post facto study
4. How can the low return rate of mail surveys be improved?
5. Why is the use of personal interviews declining?
6. Distinguish between achievement and aptitude tests.
7. _____ are used to assess a specific attribute or ability.
 - a. Surveys
 - b. Questionnaires
 - c. Pilot studies
 - d. Inventories

8. The general group of interest is the _____. The group that is selected to represent the general group is a _____.
9. A test can be _____ and not be _____.
 - a. valid; accurate
 - b. reliable; valid
 - c. valid; reliable
 - d. split; halved
10. What is random sampling? With replacement? Without replacement?
11. What is stratified random sampling, and why is it used?
12. Distinguish among the single-strata, cross-sectional, and longitudinal approaches to research.

■ Key Terms

Descriptive research methods, 59	Pilot testing, 71	Population, 80
Naturalistic observation, 61	Demographic data, 74	Sample, 80
Reactance or reactivity effect, 62	Validity, 77	Random sample, 80
Hawthorne effect, 62	Content validity, 77	Random sampling without replacement, 80
Time sampling, 64	Interrater reliability, 77	Random sampling with replacement, 80
Situation sampling, 64	Concurrent validity, 77	Stratified random sampling, 81
Interobserver reliability, 65	Criterion validity, 77	Single-strata approach, 82
Correlational study, 66	Reliability, 77	Cross-sectional research, 82
Positive correlation, 66	Test-retest procedure, 78	Longitudinal research project, 82
Negative correlation, 66	Split-half technique, 78	Cohort, 82
Zero correlation, 67	Achievement test, 78	
Ex post facto study, 69	Aptitude test, 79	
Descriptive survey, 70	Personality test or inventory, 79	
Analytic survey, 71		

■ Looking Ahead

In Chapters 3 and Chapter 4 we have considered approaches to gathering data that do not include the direct manipulation of any variables or factors by the researcher. These approaches therefore do not qualify as true experiments. In Chapter 5 we begin our consideration of experiments. First we carefully examine the scientific method and the variables involved in an experiment; then we consider the procedures involved in the control of these variables.

Using the Scientific Method in Psychology

CHAPTER

5

Components of the Scientific Method

- Objectivity • Confirmation of Findings
- Self-Correction • Control

The Psychological Experiment

- Independent Variable • Dependent Variable
- Extraneous Variables

Establishing Cause-and-Effect Relations

Formulating the Research Hypothesis

Characteristics of the Research Hypothesis

- Types of Statements • Types of Reasoning
- A New View of Hypothesis Testing
- Directional Versus Nondirectional Research Hypotheses

In Chapter 3 you learned that in 1879 Wundt used the scientific method to gather new information of interest to the new discipline of psychology. Over 100 years later, psychologists continue to believe that the scientific approach is best suited for adding to our knowledge of psychological processes.

The key elements in the scientific approach are

1. Objective measurements of the phenomenon under consideration
2. The ability to verify or confirm the measurements made by other individuals
3. Self-correction of errors and faulty reasoning
4. Exercising control to rule out the influence of unwanted factors.

We will discuss each of these characteristics in the next section. For now, we will simply say that the scientific method attempts to provide objective information so that anyone who wishes to repeat the observation in question can verify it.

Components of the Scientific Method

In this section we describe the features that characterize the scientific method. We will have much more to say about these characteristics in subsequent chapters.

Objectivity

In conducting a research project, the psychologist, like the good detective, strives to be objective. For example, psychologists select research participants in such a manner as to avoid biasing factors (e.g., age or sex). Researchers frequently make their measurements with instruments in order to be as objective as possible. We describe such measurements as being **empirical** because they are based on objectively quantifiable observations.

Empirical Objectively quantifiable observations.

Confirmation of Findings

Because the procedures and measurements are objective, we should be able to repeat them and *confirm* the original results. Confirmation of findings is important for establishing the validity of research. Psychologists use the term **replication** to refer to a research study that is conducted in exactly the same manner as a previous study. By conducting a replication study, the scientist hopes to confirm previous findings. Other studies may constitute *replication with extension*, in which scientists generate new information at the same time as they confirm previous findings. For example, Matisha Montgomery, a student at the University of Central Oklahoma in Edmond, and her faculty advisor, Kathleen Donovan, replicated and extended the conditions reported to be responsible for the “Mozart effect” (the finding that participants’ cognitive and spatial performance was better after they listened to music composed by Mozart). Their results indicated that students actually performed worse after listening to Mozart’s music (Montgomery & Donovan, 2002).

Replication

An additional scientific study that is conducted in exactly the same manner as the original research project.

Self-Correction

Because scientific findings are open to public scrutiny and replication, any errors and faulty reasoning that become apparent should lead to a change in the conclusions we reach. For example, some early American psychologists, such as James McKeen Cattell, once believed that intelligence was directly related to the quality of one’s nervous system; the better the nervous system, the higher the intelligence (see Goodwin, 2005). To verify this predicted relation, Cattell attempted to demonstrate that college students with faster reaction times (therefore, having better nervous systems) earned higher grades in college (had higher levels of intelligence). However, his observations failed to support the predicted relation, and Cattell changed his view of intelligence and how to measure it.

Control

Probably no single term characterizes science better than **control**. Scientists go to great lengths to make sure their conclusions accurately reflect the way nature operates.

Imagine that an industrial psychologist wants to determine whether providing new, brighter lighting will increase worker productivity. The new lighting is installed, and the industrial psychologist arrives at the plant to monitor production and determine whether productivity increases.

Control To directly manipulate (1) a factor of interest in a research study to determine its effects or (2) other, unwanted variables that could influence the results of a research project.



There is a problem with this research project that needs to be controlled. What is the nature of this problem, and how can it be corrected?

The main problem with this research project concerns the presence of the psychologist to check on production after installation of the new lighting. If the researcher was not present to observe production before the lighting changes were made, then he or she should not be present following the implementation of these changes. If production increases following the lighting changes, is the increase due to the new lighting or to the presence of the researcher who is monitoring production? Unfortunately, there is no way of knowing. The psychologist must exercise control to make sure that the only factor that could influence productivity is the change in lighting; other factors should not be allowed to exert an influence by varying also.

This example of research on the effects of lighting on worker productivity (which is an actual research study) also illustrates another use of the term *control*. In addition to accounting for the effects of unwanted factors, *control* can refer to the direct manipulation of the factor of major interest in the research project. Because the industrial psychologist was interested in the effects of lighting on productivity, a change in lighting was purposely created (control or direct manipulation of factors of major interest), whereas other, potentially influential but undesirable factors, such as the presence of the psychologist, were not allowed to change (control of unwanted factors).

When researchers implement control by directly manipulating the factor that is the central focus of their research, we say that an experiment has been performed. Because most psychologists believe that our most valid knowledge is produced by conducting an experiment, we will give this topic additional coverage in the next section. (Although the psychological experiment may produce the most valid data, as you saw in Chapters 3 and 4 there are numerous nonexperimental research methods that also can yield important data.)

Experiment An attempt to determine the cause-and-effect relations that exist in nature. Involves the manipulation of an independent variable (IV), recording of changes in a dependent variable (DV), and control of extraneous variables.

Independent variable (IV) A stimulus or aspect of the environment that the experimenter directly manipulates to determine its influences on behavior.

The Psychological Experiment

In many respects you can view an **experiment** as an attempt to determine the cause-and-effect relations that exist in nature. Researchers are interested in determining those factors that result in or cause predictable events. In its most basic form, the psychological experiment consists of three related factors: the independent variable, the dependent variable, and extraneous variables.

Independent Variable

The factor that is the major focus of the research and that the researcher directly manipulates is known as the **independent variable (IV)**: *independent* because it can be directly manipulated by the investigator, and *variable* because it is able to assume two or more values (often called *levels*). The IV is the causal part of the relation we seek to establish. Lighting, the IV in our previous example, had two values: the original level and the new, brighter level. *Manipulation of the IV corresponds to one use of the term control.*

Dependent Variable

The **dependent variable (DV)** consists of the recorded information or results (frequently called *data*; the singular is *datum*) of the experiment. The DV is the effect half of the cause-and-effect relation we are examining. In our example, level of productivity was the DV; the researcher measured it under the two conditions of the IV (original and brighter lighting). The term *dependent* is used because if the experiment is conducted properly, changes in DV scores will result from (depend on) the manipulation of the IV; the level of productivity will depend on the changes in lighting.

Dependent variable (DV)
A response or behavior that the experimenter measures. Changes in the DV should be caused by manipulation of the independent variable (IV).

Extraneous Variables

Extraneous variables are those factors, other than the IV, that can influence the DV and change the results of an experiment. Suppose an experimenter asks her participants to complete several tasks. At the conclusion of the experiment, she compares performance on the tasks and finds that there is a large performance difference among the groups. Was the difference caused by the different tasks (the IV) or by the sequence of performing the tasks (an extraneous variable)? Unfortunately, when an extraneous variable is present, we have no way of knowing whether the extraneous variable or the IV caused the effect we observe. Robyn Scali, a student at Catawba College in Salisbury, North Carolina, and her faculty advisor, Sheila Brownlow, were faced with just this problem; they asked their participants to complete three tasks. How did they deal with this potential extraneous variable? They stated, “Because any given task may have influenced performance on subsequent tasks, we presented tasks in one of six different orders, with each task appearing first, second, and third, exactly two times” (Scali & Brownlow, 2001, p. 7). Obviously, attention to extraneous variables is very important; it represents another use of the term *control*.

Extraneous variables
Undesired variables that may operate to influence the dependent variable (DV) and thus invalidate an experiment.

Establishing Cause-and-Effect Relations

Why do psychologists hold experimental research in such high regard? The answer to this question involves the type of information that we gain. Although we might be very objective in making our observations and even though these observations are repeatable, unless we have directly manipulated an IV, we cannot really learn anything about cause and effect in our research project. Only when we manipulate an IV and control potential extraneous variables are we able to infer a cause-and-effect relation.

What is so important about establishing a cause-and-effect relation? Although objective, repeatable observations can tell you about an interesting phenomenon, these observations cannot tell you why that phenomenon occurred. Only when we can give a cause-and-effect explanation do we begin to answer the *why* question.

For example, Naomi Freeman, a student at Earlham College in Richmond, Indiana, and her faculty advisor, Diana Punzo, wondered whether jurors were more persuaded by eyewitness testimony or by DNA evidence. To answer this research question, they conducted an experiment (Freeman & Punzo, 2001) in which college students served as mock jurors. The student

participants read one of two excerpts from a court transcript of a first-degree murder trial. In one excerpt the prosecution produced evidence from an eyewitness. In a second excerpt the prosecution produced DNA evidence. A practicing attorney assisted in the preparation of the excerpts. After the participants completed reading their particular excerpt, they indicated whether they thought the defendant was guilty.

The IV (or *cause* in the cause-and-effect relation that they were examining) was the type of evidence the prosecution presented in the transcript: eyewitness testimony or DNA evidence. At the outset of the experiment, Freeman and Punzo hypothesized that the eyewitness testimony condition would result in more guilty verdicts than the DNA condition. Why? On the basis of previous research, they believed that “although DNA evidence is extremely reliable, it is often ignored by jurors” (p. 110).



Review the experiment we have just described. What extraneous variables did the researchers control? Should they have controlled any other extraneous variables? What was their DV?

Freeman and Punzo used several controls when they conducted this research. First, they randomly determined which transcript a participant read. Second, in order to create an equivalent sense of reality for the court transcript excerpts, they asked an attorney to assist in their preparation. The type of evidence should be the only difference between the excerpts.

What other controls have you thought of? How about what happens to the excerpt when a participant finishes reading it? Should the participant be allowed to keep it? Unlike the case in a real courtroom, if the excerpt is retained, the participant could refer to it when making a decision of guilty or not guilty. To control for this potential problem, all participants returned their excerpt before answering any questions. Even though they tested the participants in small groups as they completed the experiment, the experimenters made sure that each participant answered all questions independently; collaboration with other participants would nullify our confidence in the cause-and-effect relation the researchers were trying to establish.

The participants' guilty or not guilty verdict was the DV. By comparing the number of guilty verdicts between the groups, Freeman and Punzo (2001) hoped to show that the type of evidence (cause) resulted in differences in guilty verdicts (effect); recall that they believed there would be more guilty verdicts from the participants who read the eyewitness excerpt.

Just as you would at the end of a detective case, you now want to know what happened. Contrary to the initial predictions, the results showed that participants who read the DNA transcript produced significantly more guilty verdicts than did the participants who read the eyewitness transcript. By controlling extraneous variables, Freeman and Punzo (2001) established a cause-and-effect relation between type of evidence presented to mock jurors (IV) and number of guilty verdicts (DV). The greater our control of extraneous variables, the clearer our view of the cause-and-effect relation between the IV and the DV becomes. This arrangement of IV, DV, and control of extraneous variables in Freeman and Punzo's experiment is diagrammed in Figure 5-1.

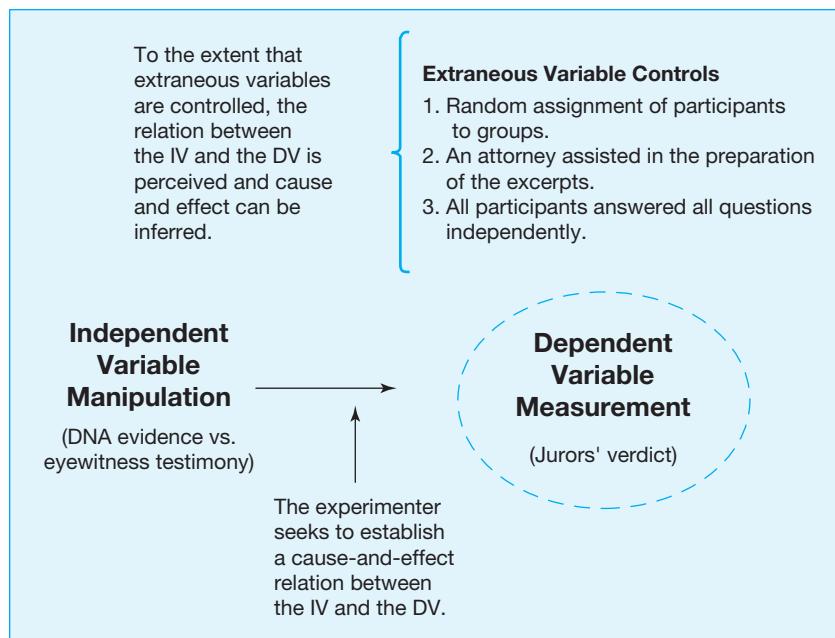


FIGURE 5-1 Diagram of the Relation of the IV, DV, and Extraneous Variable Controls in an Experiment on the Effect of DNA Evidence and Eyewitness Testimony on Jurors' Verdicts.

As we saw in Chapter 1, the first step in the research process is identifying a problem to be investigated. Then you review the literature to determine what research has been conducted in this area. Once you have completed your literature review, your own research begins to take shape. The next task is to develop a *research hypothesis*—a formal statement of your research question—taking into account what you learned from searching the literature.

Formulating the Research Hypothesis

Recall from Chapter 1 that the hypothesis is an attempt to organize data and IV–DV relations within a specific portion of a larger, more comprehensive research area or theory. Thus, hypotheses that have been supported by experimental research can make important contributions to our knowledge base.

Because you have yet to conduct your research project, the **research hypothesis** or **experimental hypothesis** is your *prediction* about the relation that exists between the IV that you are going to manipulate and the DV that you will record. If the results of your experiment support the research hypothesis, then it has the potential to make a contribution to theory; you have some grounds on which to infer a cause-and-effect relation. In order to understand the nature of the research hypothesis, we will examine some of its general characteristics.

Research or experimental hypothesis The experimenter's predicted outcome of a research project.

Characteristics of the Research Hypothesis

For the detective and the psychologist, all acceptable research hypotheses share certain characteristics: They are stated in a certain manner, they involve a certain type of reasoning, and they are presented in a certain format.

Types of Statements

Because a research hypothesis is nothing more than a statement of what you believe will occur when you conduct an experiment, you must carefully consider the statements used in constructing it.

Synthetic statement

Statements that can be either true or false.

Analytic statement

Statements that are always true.

Contradictory statement

Statements that are always false.

Synthetic, Analytic, and Contradictory Statements A statement can be one of three types: synthetic, analytic, or contradictory.

Synthetic statements are those statements that can be *either true or false*. The statement “Abused children have lower self-esteem” is synthetic because, although there is a chance that it is true, there also is a chance that it is false.

Analytic statements are those statements that are *always true*. For example, “I am making an A or I am not making an A” is an analytic statement; it is always true. You are either making an A or you are not making an A; no other possibilities exist.

Contradictory statements are those statements that are *always false*. For example, “I am making an A and I am not making an A” is a contradictory statement; it is always false. You cannot make an A and not make an A at the same time.



Review the three types of statements presented above. Which one is best suited for use in a research hypothesis?

General implication form

Statement of the research hypothesis in an “if . . . then” form.

When you conduct an experiment, you are attempting to establish the existence of a cause-and-effect relation. At the outset of the experiment you do not know whether your prediction is correct or not. Therefore, synthetic statements, which can be true or false, must constitute the research hypothesis. If your research hypothesis is an analytic or contradictory statement, there is no need (or way) to conduct research on that topic. You already know what the outcome will be merely by reading the statements.

General Implication Form As we mentioned previously, you must be able to state the research hypothesis in **general implication form** (“if . . . then” form). The “if” portion of such a statement refers to the IV manipulation you are

going to make, and the “then” portion refers to the DV changes you expect to observe. An example of a general implication form statement would be the following:

If students in one group of third-graders receive an M&M® each time they spell a word correctly, then their spelling performance will be better than that of a group of third-graders who do not receive an M&M for each correctly spelled word.

If you have read articles in psychological journals, you are probably saying to yourself, “I don’t recall seeing many statements in general implication form.” You are probably correct; most researchers do not formally state their research hypothesis in strict general implication form. For example, the hypothesis about the third-graders and their spelling performance might have been stated like this:

Third-graders who receive an M&M each time they spell a word correctly will spell better than third-graders who do not receive an M&M for each correctly spelled word.

Regardless of how the research hypothesis is stated, it must be restatable in general implication form. If it cannot be stated in this manner, there is a problem with either your IV manipulation or the DV you have chosen to measure.



Starting with “If students in one group of third-graders . . . ,” read the last general implication statement again. What IV is the researcher manipulating? What is the DV? In addition to being stated in general implication form, is this statement synthetic, analytic, or contradictory?

Whether the students receive an M&M after correctly spelling a word is the IV. The spelling performance of the two groups of third-graders is the DV. Because this statement has the potential to be true or false, it is synthetic.

The use of synthetic statements presented in general implication form highlights two additional, but related, characteristics of the research hypothesis. The **principle of falsifiability**, the first characteristic, means that when an experiment does not turn out as you predicted, this result is seen as evidence that your hypothesis is false. If, after the experiment, the two groups of third-graders do not differ in spelling ability, then you must conclude you made a bad prediction; your research hypothesis was not an accurate portrayal of nature. Even though you do not want such results to occur, your research must be capable of producing results that do not support your experimental hypothesis.

Because you use a synthetic statement for your research hypothesis, your results will never prove its truth absolutely; this is the second characteristic of the research hypothesis. Assume you did find that third-graders who received M&Ms spelled better than third-graders who did not receive M&Ms. Later you decide to replicate your experiment. Just because you obtained positive results the first time you conducted the experiment does not prove that your research hypothesis is unquestionably true and that you will always obtain positive results. When you conduct a replication, or any experiment for that matter, your hypothesis contains a synthetic statement that is either true or false. Thus, you can never absolutely prove a research hypothesis; you simply cannot yet disprove it. Certainly, as the number of

Principle of falsifiability
Results not in accord with the research hypothesis are taken as evidence that this hypothesis is false.

experiments that support the research hypothesis increases, your confidence in the research hypothesis increases.

Types of Reasoning

In stating your research hypothesis, you also must be aware of the type of reasoning or logic you use. As we will see, inductive and deductive reasoning involve different processes.

Inductive logic

Reasoning that proceeds from specific cases to general conclusions or theories.

Inductive Logic **Inductive logic** involves reasoning from specific cases to general principles. Inductive logic is the process that is involved in the construction of theories; the results of several independent experiments are considered simultaneously, and general theoretical principles designed to account for the behavior in question are derived.

For example, John Darley and Bibb Latané (1968) were intrigued by a famous incident that occurred in 1964 in the Queens section of New York: A young woman named Kitty Genovese was stabbed to death. Given the number of murders that take place each year in any large city, this event may not seem especially noteworthy. An especially horrifying aspect of this murder, however, was the fact that the killer attacked the young woman *three separate times* in the course of half an hour and that 38 people witnessed the attacks or heard the young woman's screams. The killer was frightened off twice when people turned their lights on or called from their windows; however, both times he resumed his attack. None of the people who witnessed the attack came to the victim's aid, and no one called the police while she was being attacked. Why?

Darley and Latané reported the results of several experiments they conducted. On the basis of their results, they theorized that individuals are more likely to give assistance if they are alone than if others are present. The finding that groups of bystanders are less likely than individuals to aid a person in trouble is known as the *bystander effect*. The development of this principle is an example of the use of inductive reasoning; several specific results were combined to formulate a more general principle.

Deductive logic

Reasoning that proceeds from general theories to specific cases.

Deductive Logic **Deductive logic** is the converse of inductive logic; we reason from general principles to specific conclusions or predictions. Deductive logic is the reasoning process we use in formulating our research hypothesis. By conducting a search of the literature, we have assembled a large amount of data and considered several theories. From this *general* pool of information we seek to develop our research hypothesis, a statement about the relation between a *specific* IV and a *specific* DV. For example, on the basis of previous research conducted on the bystander effect, a social psychologist might make the following deductive statement:

If a person pretends to have a seizure on a subway, that person will receive less assistance when there are more bystanders present.



Read the preceding statement once again. Is this statement acceptable as a research hypothesis?

Yes, this statement would appear to be acceptable as a research hypothesis. It is a synthetic statement (it can be true or false) presented in general implication form ("if . . . then"). Moreover, deductive logic is involved: We have gone from a general body of knowledge (about the bystander effect) to make a specific prediction concerning a specific IV (number of bystanders present) and a specific DV (receiving assistance).

Certainly we are not maintaining that deductive and inductive reasoning are totally separate processes. They can, and do, interact with each other. As you can see from Figure 5-2, the results of several initial experiments in a particular research area may lead to the development of a theory in that area (inductive reasoning).

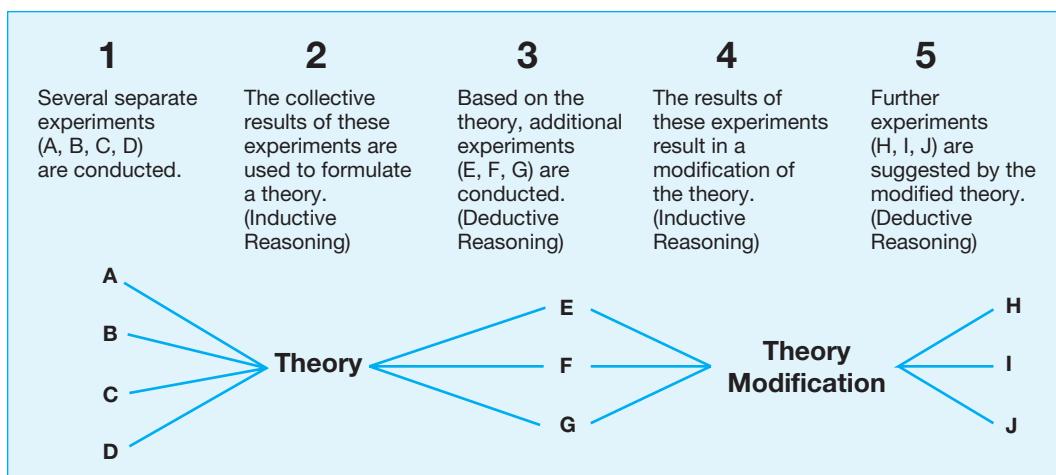


FIGURE 5-2 The Relations Between Deductive and Inductive Reasoning.

Source: Wann, D. L. & Dolan, T. J. (1994). "Spectator's evaluations of rival and fellow fans." *The Psychological Record*, 44, 351–358.

In turn, an examination of that theory and past research may suggest a specific research project (deductive reasoning) that needs to be conducted. The results of that project may result in modification of the theory (inductive reasoning), which prompts the idea for several additional research projects (deductive reasoning), and so forth. Clearly, the research process we described in Chapter 1 is based on the interplay between these two types of logical reasoning.

A New View of Hypothesis Testing

Robert Proctor and John Capaldi, psychologists at Purdue University, have presented a new view of hypothesis testing that warrants considerable thought. They based their comments on the premise that research methodology is not a static, finished process; it is constantly in a state of change and flux. Principles may be rejected or modified. The data that researchers gather really determine the methods that are accepted and continue to be used.

They contend that hypothesis testing is not a simple affair; it "is a great deal more complicated than it seems on the surface" (Proctor & Capaldi, 2001, p. 179). Several factors can determine what the researcher does when an experimental hypothesis is confirmed (or disconfirmed). For example, the *importance* of the research problem may determine whether a

researcher sticks with the problem, even if a specific hypothesis is disconfirmed. Likewise, the *promise* of the research program to answer theoretical questions may determine whether that research program flourishes.

Early in the developmental stages of a theory, hypothesis testing actually may be harmful. Because the researchers do not know all of the relevant variables (IVs, DVs, and, especially, extraneous variables), it may be very easy to disconfirm an experimental hypothesis and thereby disconfirm the theory. Proctor and Capaldi also argue that researchers are never really able to achieve a clear-cut test of a hypothesis. Why? No hypothesis is ever tested in complete isolation; we are always testing other things at the same time. For example, the adequacy of our research equipment is tested at the same time as we test a hypothesis. If the hypothesis was disconfirmed, was it due to an incorrect hypothesis or inadequate equipment? Likewise, we always make auxiliary assumptions, such as the effectiveness of randomization as a control technique, when we test a hypothesis. If these auxiliary assumptions are incorrect, we may reject a valid experimental hypothesis.

What should researchers do? Proctor and Capaldi suggest that researchers use more inductive logic when a research area is new, because there is a high probability of disconfirming hypotheses when they might be true. Their suggestion is to let the data be your guide. Sherlock Holmes made this point quite eloquently when he said, "I make a point of never having any prejudices, and of following docilely wherever fact may lead me" (Doyle, 1927, p. 407). The empirical results of your research will help reveal important relations. Hypothesis testing is important and valuable when it works; however, researchers must be sensitive to the drawbacks associated with it.

Directional Versus Nondirectional Research Hypotheses

Directional research hypothesis Prediction of the specific outcome of an experiment.

Finally, we must consider whether we are going to predict the direction of the outcome of our experiment in our research hypothesis. In a **directional research hypothesis** we specify the outcome of the experiment. For example, if we test two groups, we could entertain *one* of the following directional hypotheses:

Group A will score significantly higher than Group B.

or

Group B will score significantly higher than Group A.

In either case, we are directly specifying the direction we predict for the outcome of our experiment. (Note that although we can entertain either of these directional hypotheses, we cannot consider both of them at the same time.)

Nondirectional research hypothesis A specific prediction concerning the outcome of an experiment is not made.

On the other hand, a **nondirectional research hypothesis** does not predict the exact directional outcome of an experiment; it simply predicts that the groups we are testing will differ. Using our two-group example once again, a nondirectional hypothesis would indicate that

Group A's scores will differ significantly from Group B's.

For this hypothesis to be supported, Group A can score either significantly higher or significantly lower than Group B.



Review the differences between directional and nondirectional hypotheses. How would you write a directional research hypothesis for the M&Ms and third-grade spelling experiment we described earlier? How would you write a nondirectional hypothesis for this same experiment?

A directional hypothesis might read as follows:

If students in one group of third-graders receive an M&M each time they spell a word correctly, then their spelling performance *will be better* than that of a group of third-graders who do not receive an M&M for each correctly spelled word.

Of course, you might predict that receiving M&Ms would cause spelling performance to decrease and fall below that of the group that did not receive the M&Ms. In either instance, you would be using a directional hypothesis.

A nondirectional hypothesis would read as follows:

If students in one group of third-graders receive an M&M each time they spell a word correctly, then their spelling performance *will differ* from that of a group of third-graders who do not receive an M&M for each correctly spelled word.

In this instance, we are simply predicting that the two groups will perform differently.

Which type of hypothesis, directional or nondirectional, should you choose? If the theory you are testing calls for it, and you are *relatively certain* of your prediction, then you may want to use a directional hypothesis. For reasons we will discuss later, your chances of finding a statistically significant result are increased when you use a directional hypothesis. However, if you adopt a directional hypothesis, there is no changing your mind. If the results turn out exactly opposite to your prediction, the only thing you can say is that you were wrong and that nature doesn't operate as you thought it did. Because nature has a way of playing cruel tricks on our predictions, most researchers take a more conservative approach and state a nondirectional research hypothesis. Even though it may be slightly more difficult to achieve statistical significance, there is less potential for disappointment with the outcome of the research project when you use a nondirectional hypothesis.

■ REVIEW SUMMARY

1. Three characteristics of the scientific method are the use of (a) objective (**empirical**) findings that can be (b) confirmed by others and (c) corrected, if necessary, by subsequent research.
2. **Control** also is a distinguishing characteristic of science. Control can refer to (a) procedures for dealing with undesired factors in an experiment and/or (b) manipulation of the factor of main interest in an experiment.
3. Experimenters seek to establish cause-and-effect relations between variables they manipulate (**independent variables** or **IVs**) and behavioral changes (**dependent variables** or **DVs**) that result from those manipulations. Control also is exercised over **extraneous variables** (unwanted factors) that can influence the dependent variable.

4. Once the literature review is complete, the researcher is ready to develop a **research** or **experimental hypothesis**—the predicted outcome of the experiment.
5. The research hypothesis contains **synthetic statements** that can be either true or false and is stated in **general implication form** ("if . . . then").
6. The **principle of falsifiability** indicates that, when an experiment does not turn out as predicted, the truthfulness of the research hypothesis is discredited. Because a research hypothesis is composed of synthetic statements, it can never be absolutely proved; it can only be disproved.
7. The development of the research hypothesis involves the use of **deductive logic** in which we reason from general principles to specific cases. When we reason from specific cases to general principles, we use **inductive logic**.
8. A **directional research hypothesis** is used when the *specific* direction or nature of the outcome of an experiment is predicted. When the experimenter does not predict the specific nature of the group differences, only that they *will differ*, a **nondirectional research hypothesis** is used.

Check Your Progress

1. What are (describe) the components of the scientific method?
2. Explain what is meant by the "self-correcting nature of science."
3. A researcher conducts a study to confirm the effects of a previously tested drug. This is an example of
 - a. objectivity
 - b. replication
 - c. control
 - d. repeated measures
4. Explain the nature of the cause-and-effect relations that psychological research attempts to establish.
5. Leo conducts a study that examines the effects of extracurricular involvement on levels of self-esteem. In this study, extracurricular involvement serves as the
 - a. dependent variable
 - b. independent variable
 - c. extraneous variable
 - d. intrinsic variable
6. _____ are always true.
 - a. Synthetic statements
 - b. Contradictory statements
 - c. Statements presented in general implication form
 - d. Analytic statements

7. Why are synthetic statements used in the experimental hypothesis?
8. Describe general implication form. Be thorough.
9. Which of the following is *not* associated with the principle of falsifiability?
 - a. synthetic statements
 - b. general implication form
 - c. experimental hypothesis
 - d. analytic statements
10. The construction of theories involves _____ logic; the development of the experimental hypothesis involves _____ logic.
11. Concerning research hypotheses, most researchers
 - a. state a directional research hypothesis
 - b. state a nondirectional research hypothesis
 - c. do not use a research hypothesis
 - d. use a research hypothesis that makes no prediction

■ Key Terms

Empirical, 86	Research or experimental hypothesis, 91	Inductive logic, 94
Replication, 87	Synthetic statement, 92	Deductive logic, 94
Control, 87	Analytic statement, 92	Directional research hypothesis, 96
Experiment, 88	Contradictory statement, 92	Nondirectional research hypothesis, 96
Independent variable (IV), 88	General implication form, 92	
Dependent variable (DV), 89	Principle of falsifiability, 93	
Extraneous variables, 89		

Conducting a Good Experiment I: Variables and Control

CHAPTER

6

The Nature of Variables

Operationally Defining Variables

Independent Variables

- Types of IVs

Extraneous Variables (Confounders)

Dependent Variables

- Selecting the DV

- Recording or Measuring the DV
- Recording More Than One DV
- Characteristics of a Good DV

Nuisance Variables

Controlling Extraneous Variables

- Basic Control Techniques

In this chapter we begin to examine the methods and procedures that will allow us to make cause-and-effect inferences. We begin by carefully examining the types of variables used by the researcher: independent, extraneous, dependent, and nuisance variables. Recall from Chapter 5 that the experimenter directly manipulates *independent* variables; *dependent* variables change in response to the independent variable manipulation; and *extraneous* variables can invalidate our experimental results. As we will see, *nuisance* variables cause our results to be less clear. Once we have discussed these variables and made their relation to the experiment clear, we will consider the procedures that researchers have developed to keep unwanted, extraneous variables from influencing the results of our experiment.

The Nature of Variables

Variable An event or behavior that can assume two or more values.

Before jumping into a discussion of independent variables (IV), let's look at the nature of variables in general. A **variable** is an event or behavior that can assume at least two values. For example, temperature is a variable; it can assume a wide range of values. The same could be said for height, weight, lighting conditions, the noise level in an urban area, anxiety, confidence, and your responses to a test, as well as many other possibilities; each of these events can assume two or more values or levels.

So, when we discuss variables involved in a psychological experiment, we are talking about events or behaviors that have assumed *at least two values*. If the IV has only one level, we would have nothing against which to compare its effectiveness. Assume you want to demonstrate that a new brand of toothpaste is the best on the market. You have a group of participants try the new toothpaste and then rate its effectiveness. Even though the entire group rates the toothpaste in question as "great," you still cannot claim that it is best; you do not have ratings from other groups using different brands.

Just as the IV must have at least two values, the dependent variable (DV) also must be able to assume two or more values. Your toothpaste study would be meaningless if the *only* response the participants can make is “great”; more than one response alternative is needed.

The same logic applies in the case of extraneous variables. If two or more values are not present, then the event in question is not an extraneous variable. If all the participants in our toothpaste study are women, then we do not have to be concerned with sex differences between the groups that we test. (This point will be important later in the Controlling Extraneous Variables section of this chapter.) Notice that our concern about extraneous variables is quite different from our concern about IVs and DVs. Whereas we were concerned that the IV and DV have or are able to assume two or more values, we seek to avoid those instances where extraneous variables can assume two or more values.

Operationally Defining Variables

As you will recall from Chapter 5, we suggested that replication of past research can be a valuable source of research ideas. Let’s assume you have located a piece of research that you want to replicate. You carefully read how the experiment was conducted and find that each participant received a reward following every correct response. Assume this sentence is the only information you have concerning the *reward* and *response* involved in the experiment. If you asked 10 different researchers what reward they would use and what response they would record, how many different responses would you get? With this limited and vague information, chances are good that you would get as many different answers as the number of people you asked. How valid will your replication be? If you use a totally different reward and a totally different response, have you even conducted a replication?

Problems and concerns such as these led a 1920s Harvard University physicist, Percy W. Bridgman, to propose a way to obtain clearer communication among researchers and thus achieve greater standardization and uniformity in experimental methodology (Goodwin, 2005). Bridgman’s suggestion was simple: Researchers should define their variables in terms of the operations needed to produce them (Bridgman, 1927). If you define your variables in this manner, then other scientists can replicate your research by following the definitions you have given for the variables involved; such definitions are called **operational definitions**. Operational definitions have been a cornerstone of psychological research for nearly three-quarters of a century because they allow researchers to communicate clearly and effectively with each other.

To illustrate the use of operational definitions, let’s return to the reward and response situation we described previously. If we define reward as “a 45-mg Noyes Formula A food pellet,” then other animal researchers can use the same reinforcer by ordering a supply of 45-mg Formula A pellets from the P J. Noyes Company. Likewise, if we define the response as “making a bar press in an operant conditioning chamber (Lafayette Model 81335),” then another researcher can replicate our research setup by purchasing a similar piece of equipment from the Lafayette Instrument Company.

The experimenter must be able to convey clearly such information about all the variables involved in a research project. Hence, it is crucial to give operational definitions for the IV, DVs, and extraneous variables, as well as for nuisance variables.

Operational definition
Defining the independent, dependent, and extraneous variables in terms of the operations needed to produce them.

Independent Variables

Independent variables (IVs) are those variables that the experimenter purposely manipulates. The IV constitutes the reason the research is being conducted; the experimenter is interested in determining what effect the IV has. The term *independent* is used because the IV does not depend on other variables; it stands alone. A few examples of IVs that experimenters have used in psychological research are sleep deprivation, temperature, noise level, drug type (or dosage level), removal of a portion of the brain, and psychological context. Rather than attempting to list all possible IVs, it is easier to indicate that they tend to cluster in several general categories.

Types of IVs

Physiological IV A physiological state of the participant manipulated by the experimenter.

Physiological When the participants in an experiment are subjected to conditions that alter or change their normal biological state, a **physiological IV** is being used. For example, Susan Nash (1983), a student at Emporia State University in Emporia, Kansas, obtained several pregnant rats from an animal supplier. Upon their arrival at the laboratory, she randomly assigned half the rats to receive an alcohol–water mixture during gestation; the remainder received plain tap water. She switched the alcohol-exposed mothers to plain tap water when the pups were born. Thus, some rat pups were exposed to alcohol during gestation, whereas others were not. Nash tested all the pups for alcohol preference when they were adults and found that those animals that were exposed to alcohol (the physiological IV) during gestation drank more alcohol as adults. Nash received the 1983 J. P. Guilford–Psi Chi National Undergraduate Research Award for this experiment. Just as alcohol exposure was a physiological IV in Nash's experiment, administering a new drug to determine whether it is successful in alleviating schizophrenic symptoms also represents a physiological IV.

Experience IV
Manipulation of the amount or type of training or learning.

Experience When the effects of amount or type of previous training or learning are the central focus of the research, the researcher is using an **experience IV**. A study conducted by Monica Boice, a student at Saint Joseph's University in Philadelphia, and her faculty advisor, Gary Gargano, illustrates the use of experience as an IV. Boice and Gargano (2001) studied memory for items in a list as a function of the number of related cues that were presented at the time of recall.

Some participants received zero cues, whereas other participants received eight cues. The number of cues was an experience IV. The results of this study indicated that, under some conditions, receiving eight related cues actually resulted in worse memory performance than did receiving no cues.

Stimulus or environmental IV An aspect of the environment manipulated by the experimenter.

Stimulus Some IVs fall into the category of **stimulus** or **environmental variables**. When researchers use this type of IV, they are manipulating some aspect of the environment. Kathy Walter, Sammi Ervin, and Nicole Williamson, students at Catawba College in Salisbury, North Carolina, conducted a study under the direction of Sheila Brownlow in which they used a stimulus IV (Walter, Brownlow, Ervin, & Williamson, 1998). They asked 144 college students to judge various traits of women who walked barefoot and then wore high heels. The stimulus variable was whether

the person being judged was barefoot or wore heels. The results showed that when the women wore heels, student participants judged them as less sexy and more submissive than when they were barefoot.

Participant It is common to find **participant characteristics**, such as age, sex, personality traits, or academic major, being treated *as if* they are IVs.

Participant characteristics
Aspects of the participant, such as age, sex, or personality traits, which are treated as if they were IVs.



Although many researchers may treat participant characteristics as if they are IVs, they really are not. Why not?

To be considered an IV, the behavior or event in question must be directly manipulated by the experimenter. Although experimenters can manipulate physiological, experience, and stimulus IVs, they are not able to manipulate participant characteristics directly. For this reason, experimenters do not consider them to be true IVs. The experimenter does not create the participants' sex or cause participants to be a certain age. Participant characteristics or variables are best viewed as *classification*, not manipulation, variables. The categories for participant variables are created before the experiment is conducted, and the experimenter simply assigns the participants to these categories on the basis of the characteristics they display.

Extraneous Variables (Confounding)

Extraneous variables are those factors that can have an *unintended* influence on the results of our experiment. Extraneous variables influence the difference *between* groups. Figure 6-1A shows the relation of two groups without the influence of an extraneous variable; two possible effects of an extraneous variable are shown in Figures 6-1B and 6-1C. Thus, an extraneous variable can unintentionally cause groups to move closer together (Figure 6-1B) or farther apart (Figure 6-1C).



Review Figure 6-1 and the information we have presented about extraneous variables. The effect of an extraneous variable is similar to that of another major component of the experiment. What role does an extraneous variable appear to play in an experiment? How is the presence of an extraneous variable detrimental to the experiment?

The other component of an experiment that can influence the difference between groups is the IV. Thus, an extraneous variable can affect the outcome of an experiment. Just as other likely interpretations can damage a detective's case beyond repair, the presence of an extraneous variable is devastating to research; it is not possible to attribute the results of the

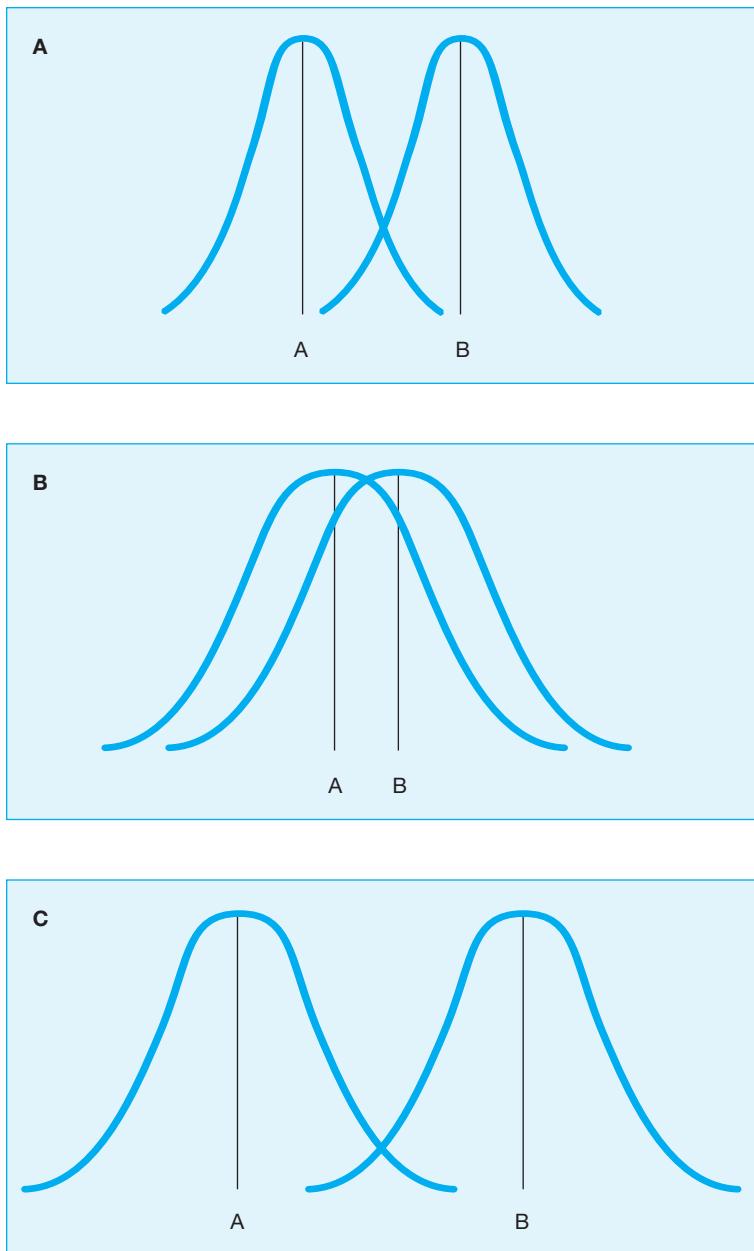


FIGURE 6-1 **A.** The difference (A = standard method; B = new method) between two groups with no confounder operating. **B.** The difference between two groups when a confounder is present and has moved the groups closer together. **C.** The difference between two groups when a confounder is present and has moved the groups farther apart.

experiment to the IV. Why? There are two variables that may have caused the groups to differ: the IV you manipulated *and* the unwanted extraneous variable. You have no way to determine which of these two variables caused the differences you observed. In such instances, when we can attribute the results either to an IV or to an extraneous variable, the experiment is **confounded**. (The terms *extraneous variable* and *confounder* are often used synonymously.) When an experiment is confounded, the best course of action is to discontinue the research and learn from your mistake. You can control the extraneous variable in the next experiment.

To illustrate how confounding works, let's consider a reading comprehension study. We have first- and second-graders available to serve as participants. The researcher assigns all the first-graders to the standard method for teaching reading comprehension and all the second-graders to the new method. The experimenter conducts the experiment and finds that the comprehension scores for students using the new method are substantially better than those of students using the standard method (see Figure 6-1C). What would have happened if the researcher had assigned the second-graders to the standard method and the first-graders to the new method? We might have seen results like those shown in Figure 6-1B. Why did these two sets of results occur? In each instance it is arguable that a preexisting difference in reading comprehension between the groups of children created differences between the two groups (i.e., the preexisting difference acted *as if* it were an IV). Assuming that second-graders have superior reading comprehension, it seems reasonable to suggest that the new method seemed even more effective when the second-graders used it (i.e., the difference between the groups was exaggerated). However, when the second-graders used the standard method, the superior method used by the first-graders increased their scores and moved the groups closer together (i.e., group differences decreased). Certainly, all of this commentary is only speculation on our part. It is also possible that the IV created the group differences that we observed. The main point is that we really do not know what caused the differences—the IV (type of method used) or the extraneous variable (grade level).

The presence of an extraneous variable is often very difficult to spot; it may take several knowledgeable individuals scrutinizing an experiment from every possible angle to determine whether one is present. If the experimenter detects an extraneous variable before conducting the research, then the experimenter can deal with the problem and proceed with the experiment. We present techniques for controlling unwanted variables later in this chapter.

Confounding A situation in which the results of an experiment can be attributed to either the operation of an IV or an extraneous variable.

Dependent Variables

The dependent variable (DV) changes as a function of the level of the IV experienced by the participant; therefore, the value the DV assumes *truly depends* on the IV. The DV consists of the data or results of our experiment. As with all aspects of psychological research, experimenters must give the DV appropriate consideration when they formulate an experiment. The experimenter must deal with such considerations as selecting the appropriate DV, deciding exactly which measure of the DV to use, and whether to record more than one DV.

Selecting the DV

Because psychology often is defined as the science of behavior, the DV typically consists of some type of behavior or response. When the researcher administers the IV, however, it is likely that several responses will occur. Which one should the researcher select as the DV? One answer to this question is to look carefully at the experimental hypothesis. If you have stated your hypothesis in general implication form ("if . . . then"—see Chapter 5), the "then" portion of the hypothesis will give you an idea of the general nature of your DV. For the Boice and Gargano (2001) memory study the choice was easy: "The dependent measure was the number of words correctly recalled" (p. 119).

What if your hypothesis were more general? Say that you wanted to study "spatial abilities." Where could you find information to help you choose a *specific* DV? We hope you are already a step ahead of us; our literature review (see Chapter 2) can provide valuable guidelines. If other researchers have used a particular response successfully as a DV in previous research, chances are that it will be a good choice again. Another reason for using a DV that has been used previously by researchers is that you will have a comparison for your results. Although totally different DVs may provide exciting new information, the ability to relate the results of experiments using different responses is more difficult.

Recording or Measuring the DV

After you have selected the DV, you will have to decide exactly how to measure or record it. Several possibilities exist.

Correctness With this DV measure, the participant's response is either correct or incorrect. Because they counted the number of words their participants remembered, Boice and Gargano (2001) used a correctness DV.

Rate or Frequency If you were studying the lever-pressing performance of a rat or pigeon in an operant conditioning chamber (Skinner box), then your DV would likely be the rate of responding shown by the animal. The rate of responding determines how rapidly responses are made during a specified time period. You can plot your data in the form of a cumulative record with steeper slopes representing higher rates (i.e., large numbers of responses being made in shorter periods of time). Figure 6-2 shows some different rates of responding.

If you were studying the number of social interactions among children during free play at a kindergarten, you might want to record the frequency, rather than the rate, of responding. Your DV, then, would simply be the number of responses shown during a specified time period without any concern for how rapidly the participant makes them.

Degree or Amount Often researchers record the DV in terms of degree or amount. In this instance, you do not record the number or frequency of the participant's responses; rather, you typically record a single number that indicates the degree or amount. Amie McKibban and Shawn Nelson, students at Emporia State University in Emporia, Kansas, studied satisfaction with life in college students (McKibban & Nelson, 2001). Scores on a Satisfaction With Life scale measured how satisfied (i.e., degree or amount) their participants were with life.

Latency or Duration In many situations, such as studies of learning and memory, how quickly participants make a response (latency) or how long the response lasts (duration) are of particular interest. For example, Rachel Ball, Erica Kargl, J. Davis Kimpel, and Shana Siewert, students at Wheaton College in Wheaton, Illinois, were interested in the relation between a participant's mood and his or her reaction time measured as a latency DV. They found that

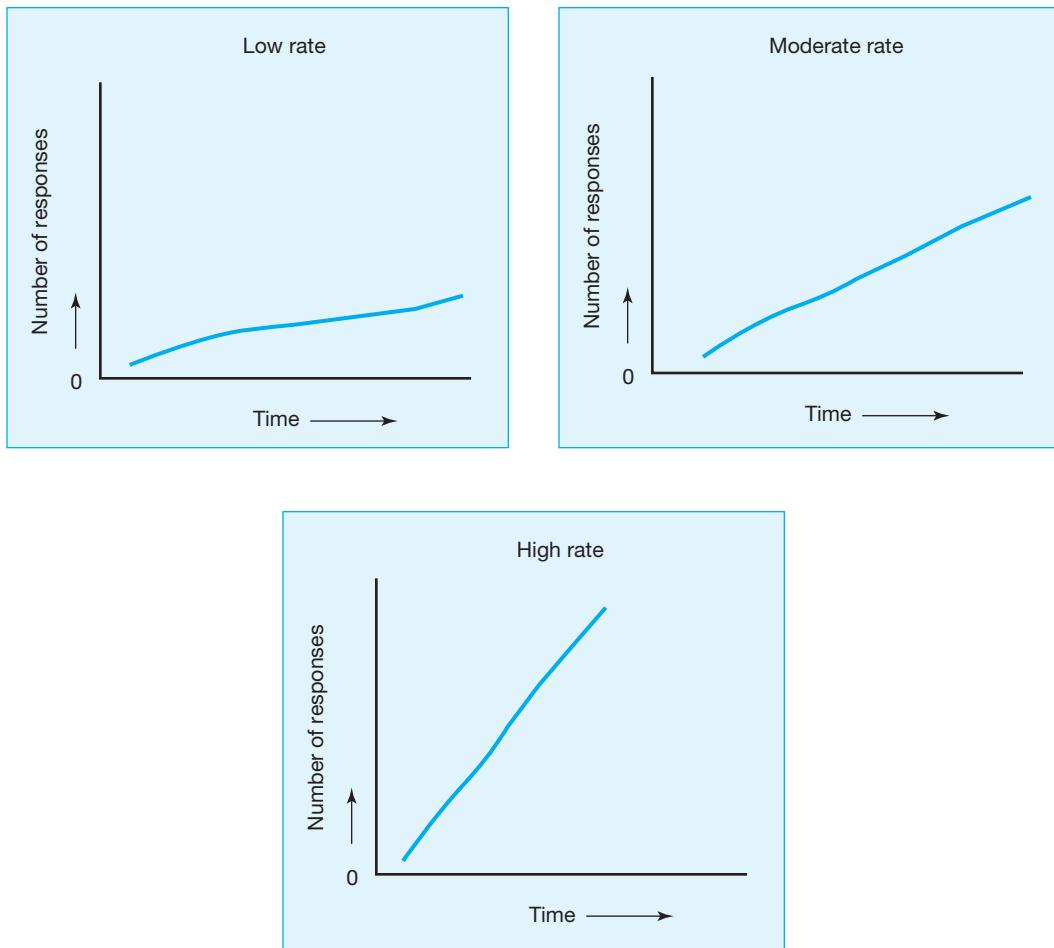


FIGURE 6-2 Different Rates of Responding.

participants in sad and suspenseful moods had longer reaction times than participants in a neutral mood (Ball, Kargl, Kimpel, & Siewert, 2001).

Recording More Than One DV

If you have the measurement capabilities, there is nothing to prohibit the recording of more than one DV. Possibly additional data will strengthen your knowledge claim in the same way it might strengthen a detective's case. Should you record additional DVs? The answer to this question really boils down to deciding whether recording additional DVs is going to add appreciably to your understanding of the phenomenon under study. If recording an additional DV makes a meaningful contribution, then you should give it serious consideration. If measuring and recording another DV does not make a substantive contribution, then it is probably not worth the added time and trouble. Often you can use previous research as a guide concerning whether you should consider recording more than one DV.



Consider a reversed eye-hand coordination (mirror tracing) experiment in which you record the time taken to complete the tracing of a star pattern (a latency DV) while looking in a mirror. Is this DV sufficient to give you a good, complete picture of the performance of this task, or should you also record another DV?

You probably should record a second DV. The latency DV indicates only how long it took to trace the star pattern. The experimenter has no record of the number of errors (going beyond the boundaries of the figure) made by the participants. A second DV, which measures the number of errors (a frequency DV), will make a significant contribution to this experiment.

The need for more than one DV was recognized in an experiment conducted by Janet Luehring, a student at Washburn University in Topeka, Kansas, and her faculty advisor, Joanne Altman (Luehring & Altman, 2000). These investigators studied male–female differences in spatial ability performance. Their participants performed a mental rotation (visualizing what an object would look like after it was rotated in space). Because the participant's performance could be correct, incorrect, or (because the task was timed) uncompleted, Luehring and Altman recorded *both correct and incorrect responses*. The number of incorrect responses, which was *not* equal to the total number of responses minus the number of correct responses, had the potential to provide additional, relevant information.

Characteristics of a Good DV

Although considerable thought may go into deciding exactly how the DV will be measured and recorded and whether more than one DV should be recorded, the experimenter still has no guarantee that a good DV has been selected. What constitutes a good DV? We want the DV to be **valid**.

The DV is **valid** when it measures what the experimental hypothesis says it should measure. For example, assume you are interested in studying intelligence as a function of the differences in regional diet. You believe the basic diet consumed by people living in different regions of the United States results in differences in intelligence. You devise a new intelligence test and set off to test your hypothesis. As the results of your project start to take shape, you notice that the scores from the Northeast are higher than those from other sections of the country; your hypothesis appears to be supported. However, a closer inspection of the results indicates that participants not living in the Northeast miss only certain questions. Are all the questions fair and unbiased, or do some favor Northerners? For example, you notice there are several questions about subways. How many individuals from Arizona are familiar with subways? Thus, your DV (scores on the intelligence test) may have a regional bias and may not measure the participants' intelligence consistently from region to region. A good DV must be directly related to the IV and must measure the effects of the IV manipulation as the experimental hypothesis predicts it will.

Valid Measuring what is supposed to be measured.

Reliable Producing consistent measurements.

A good DV is also **reliable**. If the scores on an intelligence test are used as the DV, then we would expect to see similar scores when the test is administered again under the same IV conditions (test–retest procedure; see Chapter 4). If the test gives the same individual different IQ scores at different times, it is not a reliable test.

Nuisance Variables

Nuisance variables are either characteristics of the participants or unintended influences of the experimental situation that make the effects of the IV more difficult to see or determine. It is important to understand that nuisance variables influence all groups in an experiment; their influence is not limited to one specific group. When they are present, nuisance variables result in greater variability in the DV; the scores *within* each group spread out more. For example, assume you are interested in studying reading comprehension. Can you think of a participant characteristic that might be related to reading comprehension? How about intelligence or IQ?

Figure 6-3A shows the spread of the reading comprehension scores within a group when there are not wide differences in intelligence among the participants within the group. In this instance a nuisance variable is not operating.

Nuisance variable

Unwanted variables that can cause the variability of scores within groups to increase.

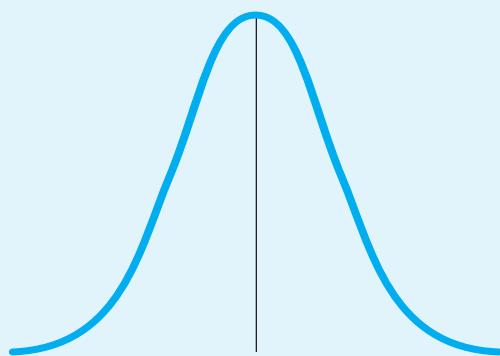
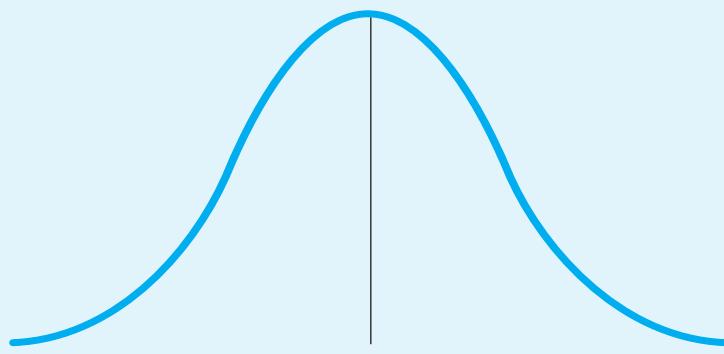
A**B**

FIGURE 6-3 **A.** The spread of scores within a group when a nuisance variable is not operating. **B.** The spread of scores within a group when a nuisance variable is operating.

You can see how the majority of the comprehension scores are similar and cluster in the middle of the distribution; there are relatively few extremely low and extremely high scores. Figure 6-3B shows the distribution of comprehension scores when there are wider differences in intelligence (i.e., a nuisance variable is present). Notice that the scores are more spread out; there are fewer scores in the middle of the distribution and more scores in the extremes.

How does a nuisance variable influence the results of an experiment? To answer that question, we need to add another group of participants to our example and conduct a simple experiment. Imagine we are evaluating two methods for teaching reading comprehension: the standard method and a new method. In Figure 6-4A, we are comparing two groups that have not been influenced by the nuisance variable. The difference between these two groups is pronounced and clear; they overlap very little.

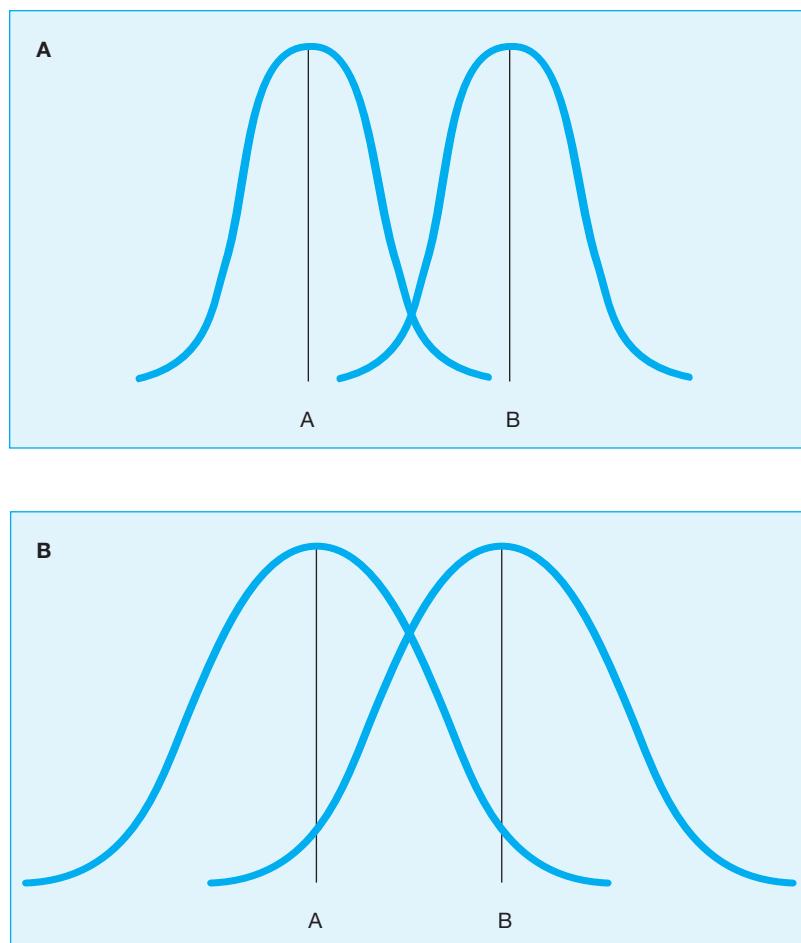


FIGURE 6-4 **A.** A Comparison of Two Groups When a Nuisance Variable Is Not Operating. **B.** A Comparison of the Same Two Groups When a Nuisance Variable Is Operating.

Let's add the effects of a nuisance variable, such as wide differences in verbal ability, to each group and then compare the groups. As you can see in Figure 6-4B, when the scores spread out more, there is greater overlap, and the difference between the groups is not as distinct and clear as when the nuisance variable was not present. When a nuisance variable is present, our view of the experimental results is clouded; we are unable to see clearly the difference the IV may have created between the groups in our experiment. Notice that when the nuisance variable was added (Figure 6-4B), the *only* thing that happened was that the scores spread out in both extremes of each distribution—the relative location of the distributions did not change. *Nuisance variables increase the spread of scores within a distribution; they do not cause a distribution to change its location.*



For each example, indicate the nuisance variable and its effect.

1. An experimenter measures reaction time in participants ranging in age from 12 to 78 years.
2. The ability of participants to recall a list of words is being studied in a room that is located by a noisy elevator.
3. The laboratory where participants are tested for manual dexterity has frequent, unpredictable changes in temperature.

In the first situation the wide age range is the nuisance variable. The younger participants should display faster reaction times than the older participants; therefore, the scores will spread into both ends of the distributions. The change in noise level caused by the operation of the elevator is the nuisance variable in the second example; the frequent, unpredictable temperature changes are the nuisance variable in the third example. The change in conditions in all three examples is likely to increase the spread of scores. Our goal as researchers is to keep nuisance variables to a minimum so that the effects of the IV are as clear as possible.

■ REVIEW SUMMARY

1. A **variable** is an event or behavior that can assume at least two levels.
2. **Independent variables** are purposely manipulated by the experimenter and form the core or purpose of the experiment. **Physiological IVs** refer to changes in the biological state of the participants, whereas **experience IVs** refer to manipulations of previous experience or learning. **Stimulus IVs** are manipulations in some aspect of the environment.
3. Although **participant characteristics** such as age, sex, or personality traits are often treated as IVs, technically they are not IVs because the experimenter does not directly manipulate them.
4. **Extraneous variables (confounders)** can have an unintended influence on the results of an experiment by changing the difference *between* the groups. When an extraneous variable is present, the experiment is confounded.
5. The **dependent variable** changes as a function of the changes in the IV. The experimental hypothesis can provide possible guidelines concerning the selection of the DV. Past research also can assist the experimenter in selecting the DV.

6. The DV can be recorded in terms of correctness, rate or frequency, degree or amount, and latency or duration. If additional information will be gained, the experimenter should consider recording more than one DV. A good DV is directly related to the IV (**valid**) and **reliable**.
7. **Nuisance variables** are variables that increase the variability of scores *within all groups*. The presence of nuisance variables makes the results of an experiment less clear.

Check Your Progress

1. An event or behavior that can assume at least two values is a _____.
2. Matching
 1. DV
 2. extraneous variable
 3. physiological IV
 4. experience IV
 5. stimulus IV
 6. participant variable
 - A. change in normal biological state
 - B. manipulation of environment
 - C. can damage the experiment and its results
 - D. age
 - E. changes as a function of changes in IV
 - F. amount of previous learning
3. Your research involves determining the effects of persuasion on the strength of attitudes. You are using a _____ measurement of the DV.
 - a. correctness
 - b. rate
 - c. degree
 - d. duration
4. A good DV has two primary qualities; it is both _____ and _____.
 - a. easy to identify; easy to measure
 - b. dependent on the experimenter; identifiable
 - c. valid; reliable
 - d. positively correlated; negatively correlated
5. Under what conditions should you record more than one DV?
6. Variables that result in greater within-group variability in the data are called
 - a. independent variables
 - b. confounders
 - c. nuisance variables
 - d. uncontrolled variables

Controlling Extraneous Variables

Just as care and precision are crucial when a detective attempts to solve a case, control forms an integral component of psychological research. The experimenter must exercise control over both extraneous variables and nuisance variables so that the results of the experiment are as meaningful (no extraneous variables present) and clear (minimal influence of nuisance variables) as possible. When you are dealing with a variable that can be clearly specified and quantified (e.g., sex, age, educational level, temperature, lighting intensity, or noise level),



Although psychologists can use many different types of experimental control in their research, none of them should be this complicated.

“It was more of a ‘triple-blind’ test. The patients didn’t know which ones were getting the real drug, the doctors didn’t know, and I’m afraid, nobody knew.”

one of the five basic control techniques should be applicable. We describe these five basic control techniques—randomization, elimination, constancy, balancing, and counterbalancing—in the next section.

Basic Control Techniques

As we discuss the basic control techniques, it is important to keep in mind that their goals are to (a) produce groups that are equivalent prior to the introduction of the IV, thereby eliminating extraneous variables, and (b) reduce the effects of nuisance variables as much as possible.

Randomization We begin our discussion of control with randomization because it is the most widely used technique. **Randomization** guarantees that each participant has an equal chance of being assigned to any group in the experiment. For example, once the students had volunteered for the research on memory, Boice and Gargano (2001) indicated that “[t]he participants were placed randomly into one of the six experimental groups” (p. 120).

The logic behind using randomization is as follows. Because all participants have an equal likelihood of being selected for each group in an experiment, any unique characteristics associated with the participants should be equally distributed across all groups that are formed. Consider level of motivation, for example. Although it may not be feasible to measure each participant’s level of motivation, this variable can still be controlled by randomly forming the groups in our experiment. Just by chance we would expect that each group would have some participants who are highly motivated, some participants who are moderately motivated, and some participants who are barely motivated. Thus, the groups should be equated with regard to the average level of motivation, as well as the myriad of other unknown and unsuspected extraneous variables that might be present.

Randomization A control technique that ensures that each participant has an equal chance of being assigned to any group in an experiment.



Even though it is the most widely used control procedure, randomization has one major drawback. What is it?

Because we are never *fully* aware of the variables that randomization controls and thus do not know whether these variables are distributed evenly to all groups, we cannot be positive that this control technique is effective. It is possible that all the highly motivated participants will be randomly assigned to the same group. Also, you need to consider what is being controlled when randomization is used. If you find yourself having some difficulty specifying exactly what is controlled by randomization, then you are on the right track. Randomization is used as a control technique for *all* variables that *might* be present and of which the experimenter is unaware. If the experimenter is unaware of exactly which variables are being controlled and how effective the control is, then it should come as no surprise that the experimenter is never *completely* sure how effective randomization has been.

Elimination When we know the extraneous variables or nuisance variables, our approach can be more direct. For example, we might choose to remove or **eliminate** the unwanted variable. This sounds easy, but in practice you may find it quite difficult to remove a variable *completely*.

Shann Sagles, Sharon Coley, Germilina Espiritu, and Patricia Zahregian, students at Pace University in White Plains, New York, and their faculty advisor, Richard Velayo, used elimination as a control in their cross-cultural study of the identification of facial expressions. “The 35-mm photos of the target individuals, taken from the base of the chin to the top of the forehead, excluded their attire and body type. The rationale for this procedure was to *eliminate* [emphasis added] extraneous variables that may have influenced the participants’ responses” (Sagles, Coley, Espiritu, Zahregian, & Velayo, 2002, p. 33). Thus, Sagles et al. guaranteed that clothing and weight did not affect their participants’ responses.

When the variable in question consists of an entire category of events, such as noise, temperature, or lighting condition, it may be difficult, if not impossible, to eliminate it. If, however, the variable is a specific occurrence within one of these more general categories, such as temperatures above 80 degrees, then it may be possible to eliminate that *aspect* of the variable. In this situation, however, the experimenter is interested not just in eliminating a variable but also in producing and maintaining a *constant* condition under which the participants in the experiment are tested.

Constancy When it is difficult or impossible to eliminate a variable completely, the experimenter may choose to exercise control by creating a uniform or constant condition which is experienced by all participants. **Constancy** has become a standard control technique for many researchers. For example, experimental testing may take place in the same room, with the same lighting and temperature levels, and at the same time of day (if the experiment is conducted on more than one day). In this instance the location of the experiment, the temperature level, the lighting level, and the time of day have not been eliminated but rather have assumed a constant value.

Elimination A control technique whereby extraneous variables are completely removed from an experiment.

Constancy A control technique by which an extraneous variable is reduced to a single value that is experienced by all participants.



Juan and Nancy are interested in determining which of two methods of teaching psychological statistics is best. Two statistics classes are available: Juan teaches one method to one class, and Nancy teaches the second method to the other class. This experiment is confounded because Juan and Nancy each teach one of the classes, so it is impossible to tell whether differences between the classes are due to the method or the teacher. How would you use constancy to control this extraneous variable?

The easiest approach would be to have only one of the experimenters, Juan or Nancy, teach both classes. Thus, the extraneous variable, the teacher, is the same for both classes and the experiment is no longer confounded.

By making sure that the experimental testing conditions do not vary unpredictably, constancy can also control for nuisance variables. When the testing conditions are the same from testing session to testing session, there is a greater likelihood that the scores within the groups will not spread out as much because of the constancy. Constancy can also control nuisance variable effects produced by participant variables such as age, sex, and educational level. For example, in a study of the effects of massage and touch on body dissatisfaction, Angela Larey, a student at Missouri Southern State College in Joplin, Missouri, “used only women [as participants] because they generally manifest greater body dissatisfaction than men” (Larey, 2001, p. 79). Recall that wide variations in a such participant variables may result in a greater spread of scores within the groups. If the technique of stratified random sampling (see Chapter 5) is used, then the variability among the participants’ scores should be smaller because the participants are more homogeneous. Clearly, this procedure helps create constancy.

Although constancy can be an effective control technique, there are situations in which the unwanted variable(s) cannot be reduced to a single value that all participants experienced in the experiment. What can be done when this variable assumes two or more values? The answer may lie in the control technique known as balancing.

Balancing **Balancing** represents a logical extension of control through constancy. Thus, the groups in our experiment are balanced or equivalent when each group experiences *all* unwanted variables or *levels* of unwanted variables in the same manner or to the same degree.

In the simplest example of balancing we would test two groups—one group (the experimental group) would receive the IV; the second group (the control or comparison group) would be treated identically but would not receive the IV. If the groups were balanced or equated with regard to extraneous variables, then we could tentatively conclude that differences between them were caused by the IV. This general situation is diagrammed in Table 6-1.

When the potential extraneous variables, such as various personality differences in human participants, are *unknown*, the experimenter uses randomization to form equivalent groups, and we assume that the respective extraneous variables are distributed equally to all groups. When the extraneous variables, such as sex of the experimenter, are known, then the experimenter can be more systematic in the use of the balancing technique to produce equivalent conditions.

Balancing A control procedure that achieves group equality by distributing extraneous variables equally to all groups.

TABLE 6-1 Balanced Extraneous Variables

When the extraneous variables are experienced in the same manner, the groups are said to be balanced.

Group 1	Group 2
Treatment A	Treatment B
Ext. Var. 1	Ext. Var. 1
Ext. Var. 2	Ext. Var. 2
Ext. Var. 3	Ext. Var. 3
Ext. Var. 4	Ext. Var. 4



Let's assume that all the students included in Juan and Nancy's experiment cannot be taught by one teacher; two are needed. How could balancing be used to remove the potential confounding caused by a different teacher teaching each of the new methods?

As you can see in Table 6-2, an easy solution to their problem is to have Juan and Nancy each teach half the students under *each* method. Thus, the two teachers appear equally under each teaching method and the classes are balanced with regard to that potential confounding variable. This teaching example illustrates the simplest situation in which balancing is used to control one extraneous variable. Balancing can also be used with several extraneous variables (see Table 6-1); the only requirement is that each extraneous variable appear equally in each group.

Although elimination, constancy, and balancing offer the experimenter powerful control techniques, they are not able to deal with all control problems. In the next section we examine one of these problem situations, *sequencing* or *order*, and how to control it through counterbalancing.

TABLE 6-2 Using Balancing to Eliminate Confounding in Teaching Two Methods of Psychological Statistics

Juan	Nancy
25 students → Method 1	25 students → Method 1
25 students → Method 2	25 students → Method 2

Counterbalancing In some experiments, participants participate in more than one experimental condition. For example, you might want to conduct a cola taste test to determine which of two brands of cola is preferred. As you set up your tasting booth at the local mall, you are sure you have taken all the right precautions: The tasting cups are all the same, the two colas will be poured from identical containers, the participants will consume the same amount of Cola A and then Cola B (in that order), and the participants will be blindfolded during the test so color

differences will not influence their choice. Your control seems to be perfect. Is it? *Constancy* is achieved by ensuring that (a) the tasting cups are the same, (b) the colas are poured from similar containers, and (c) all participants consume the same amount of each cola. By blindfolding the participants, you have *eliminated* any problems that may be caused by differences in the visual appearance of the two colas. These are all relevant control procedures.



Review the experiment we have just described. What control procedures are being used? A problem that has been overlooked needs to be controlled. What is this problem and how might it be controlled?

The problem that is overlooked concerns the sequence or order for sampling the two colas. If Cola A is *always* sampled before Cola B and one cola is liked more, you cannot be sure the preference is due to its great flavor or the fact that the colas were always sampled in the same order and this order may have influenced the participants' reactions. The technique used when a sequence or order effect must be controlled is known as **counterbalancing**. There are two types of counterbalancing: within-subject and within-group. **Within-subject counterbalancing** attempts to control the sequence effect for each participant, whereas **within-group counterbalancing** attempts to control this problem by presenting different sequences to different participants.

Within-Subject Counterbalancing Returning to the problem of sequence in our cola challenge, we could deal with this problem by having each participant sample the two colas in the following sequence: ABBA. By using within-subject counterbalancing, each participant will taste Cola A once before and once after tasting Cola B. Thus, the experience of having tasted Cola A first is counterbalanced by tasting Cola A last.

Although it may seem relatively easy to implement within-subject counterbalancing, there is one major drawback to its use: Each participant must experience each condition more than once. In some situations the experimenter may not want or be able to present the treatments more than once to each participant. For example, you may not have sufficient time to conduct your cola challenge and allow each participant the opportunity to sample *each* brand of cola more than once. In such instances within-group counterbalancing may offer a better control alternative.

Within-Group Counterbalancing Another way to deal with the cola challenge sequencing problem is randomly to assign half the participants to experience the two colas in the Cola A → Cola B sequence and the remaining half of the participants to receive the Cola B → Cola A sequence. The preference of participants who tasted Cola A before Cola B could be compared with the preference of the participants who tasted Cola B first.

Assuming that we tested six participants, the within-group counterbalanced presentation of the two colas would be diagrammed as shown in Table 6-3.

Counterbalancing A procedure for controlling order effects by presenting different treatment sequences.

Within-subject counterbalancing Presentation of different treatment sequences to the same participant.

Within-group counterbalancing Presentation of different treatment sequences to different participants.

TABLE 6-3 Within-Group Counterbalancing for the Two-Cola Challenge When Six Participants Are Tested

	Tasting 1	Tasting 2
Participant 1	A	B
Participant 2	A	B
Participant 3	A	B
Participant 4	B	A
Participant 5	B	A
Participant 6	B	A

As you can see, three participants receive the A → B sequence, whereas three participants receive the B → A sequence. This basic diagram illustrates the three requirements of within-subject counterbalancing:

1. Each treatment must be presented to each participant an equal number of times. In this example, each participant tastes Cola A once and Cola B once.
2. Each treatment must occur an equal number of times at each testing or practice session. Cola A is sampled three times at Tasting 1 and three times at Tasting 2.
3. Each treatment must precede and follow each of the other treatments an equal number of times.

In this example, Cola A is tasted first three times and is tasted second three times.

Counterbalancing is not limited to two-treatment sequences. For example, let's assume that your cola challenge involves three colas instead of two. The within-group counterbalancing needed in that situation is diagrammed in Table 6-4. Carefully examine this diagram. Does it satisfy the requirements for counterbalancing?

It appears that all the requirements have been met. Each cola is tasted an equal number of times (6), is tasted an equal number of times (2) at each tasting session, and precedes and follows

TABLE 6-4 Within-Group Counterbalancing for the Three-Cola Challenge When Six Participants Are Tested

	Tasting Session		
	1	2	3
Participant 1	A	B	C
Participant 2	A	C	B
Participant 3	B	A	C
Participant 4	B	C	A
Participant 5	C	A	B
Participant 6	C	B	A

each of the other colas an equal number of times (2). It is important to note that, if we want to test more participants, they would have to be added in multiples of 6. The addition of any other number of participants violates the rules of counterbalancing by creating a situation in which one cola is tasted more than the others, appears more often at one tasting session than the others, and does not precede and follow the other colas an equal number of times. Try adding only 1 or 2 participants to Table 6-4 to see whether you can satisfy the requirements for counterbalancing.

Table 6-4 illustrates another consideration that must be taken into account when counterbalancing is used. When only a few treatments are used, the number of different sequences that will have to be administered remains relatively small and counterbalancing is manageable. When we tested 2 colas, only 2 sequences were involved (see Table 6-3); however, the addition of only 1 more cola resulted in the addition of 4 sequences (see Table 6-4). If we added an additional cola to our challenge (Colas A, B, C, and D), we would now have to administer a total of 24 different sequences, and our experiment would be much more complex to conduct. Our minimum number of participants would be 24, and if we wanted to test more than 1 participant per sequence, participants would have to be added in multiples of 24.

How do you know how many sequences will be required? Do you have to write down all the possible sequences to find out how many there are? No. You can calculate the total number of sequences by using the formula $n!$ (n factorial). All that is required is to take the number of treatments (n), factor or break that number down into its component parts, and then multiply these factors or components. For example,

$$2! \text{ would be } 2 \times 1 = 2$$

$$3! \text{ would be } 3 \times 2 \times 1 = 6$$

$$4! \text{ would be } 4 \times 3 \times 2 \times 1 = 24$$

and so forth. When you can administer all possible sequences, you are using **complete counterbalancing**. Although complete counterbalancing offers the best control for sequence or order effects, often it cannot be attained when several treatments are included in the experiment. As we just saw, the use of 4 colas would require a minimum of 24 participants ($4! = 4 \times 3 \times 2 \times 1 = 24$) for complete counterbalancing. Testing 5 colas would increase the number of sequences (and the minimum number of participants) to 120 ($5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$). In situations requiring a large number of participants to implement complete counterbalancing, either you can reduce the number of treatments until your time, financial, and participant resources allow complete counterbalancing, or you can complete the experiment without completely counterbalancing.

Complete counter-balancing All possible treatment sequences are presented.

Incomplete counterbalancing refers to the use of some, but not all, of the possible sequences. Which sequences are to be used, and which ones are to be excluded? Some experimenters randomly select the sequences they will employ. As soon as the number of participants to be tested has been determined, the experimenter randomly selects an equal number of sequences. For example, Table 6-5 illustrates a possible random selection of sequences for conducting the cola challenge with four colas and only 12 participants.

Incomplete counter-balancing Only a portion of all possible sequences are presented.

Although random selection appears to be an easy approach to the use of incomplete counterbalancing, there is a problem. If you examine Table 6-5 carefully, you will see that although each participant receives each treatment an equal number of times, the other

TABLE 6-5

Incomplete Counterbalancing Using Randomly Selected Tasting Sequences for the Four-Cola Challenge Using 12 Participants

	Testing Session			
	1	2	3	4
Participant 1	A	B	C	D
Participant 2	A	B	D	C
Participant 3	A	C	D	B
Participant 4	A	D	C	B
Participant 5	B	A	C	D
Participant 6	B	C	D	A
Participant 7	B	C	A	D
Participant 8	C	A	B	D
Participant 9	C	B	A	D
Participant 10	C	D	B	A
Participant 11	D	B	C	A
Participant 12	D	C	B	A

requirements for counterbalancing are not satisfied. Each treatment does not appear an equal number of times at each testing or practice session, and each treatment does not precede and follow each of the other treatments an equal number of times.

Two approaches can be adopted to resolve this problem, although neither one is completely satisfactory. We could randomly determine the treatment sequence for the *first* participant and then systematically rotate the sequence for the remaining participants. This approach is diagrammed in Table 6-6.

Thus, the first participant would taste the colas in the order B, D, A, C, whereas the second participant would experience them in the order D, A, C, B. We would continue systematically rotating the sequence until each cola appears once in each row and each column. To test a

TABLE 6-6

An Incomplete Counterbalancing Approach

This approach involves randomly determining the sequence for the first participant and then systematically rotating the treatments for the following sequences.

	Tasting Sequence			
	1	2	3	4
Participant 1	B	D	A	C
Participant 2	D	A	C	B
Participant 3	A	C	B	D
Participant 4	C	B	D	A

total of 12 participants, we would assign 3 participants to each of the 4 sequences. By ensuring that each treatment appears an equal number of times at each testing session, this approach comes close to satisfying the conditions for counterbalancing. It does not, however, ensure that the treatments precede and follow each other an equal number of times. A more complex procedure, the Latin square technique, is used to address this issue. Because of its complexity, this procedure is seldom used. If you are interested in reading about its use, Rosenthal and Rosnow (1991) offer a nice presentation of its particulars.

Now that we have examined the mechanics involved in implementing complete and incomplete counterbalancing, let's see exactly what counterbalancing can and cannot control. To say simply that counterbalancing controls for sequence or order effects does not tell the entire story. Although counterbalancing controls for sequence or order effects, as you will see, it also controls for *carryover effects*. It cannot, however, control for *differential carryover*.

Sequence or Order Effects Sequence or order effects are produced by the participant's being exposed to the sequential presentation of the treatments. For example, assume we are testing reaction time to three types of dashboard warning lights: red (R), green (G), and flashing white (FW). As soon as the warning light comes on, the participant is to turn off the engine. To counterbalance this experiment completely would require 6 sequences and at least 6 participants ($3! = 3 \times 2 \times 1 = 6$). If we found the reaction time to the first warning light, regardless of type, was 10 seconds, and that increases in reaction time of 4 and 3 seconds were made to the second and third lights (regardless of type), respectively, we would be dealing with a sequence or order effect. This example is diagrammed in Table 6-7. As you can see, the sequence or order effect depends on where in

Sequence or order effects
The position of a treatment in a series determines, in part, the participants' response.

TABLE 6-7 Example of Sequence or Order Effects in a Counterbalanced Experiment

The performance decrease shown in parentheses below each sequence indicates the effect of testing reaction time to red (R), green (G), and flashing white (FW) lights on an instrument panel at that particular point in the sequence. Thus, second and third testings result in increases (i.e., slower reaction times) of 4 and 3, respectively, regardless of the experimental task.

Order of Task Presentation		
Performance Increase →	R	G
(0)	4	3)
R	FW	G
(0)	4	3)
G	R	FW
(0)	4	3)
G	FW	R
(0)	4	3)
FW	R	G
(0)	4	3)
FW	G	R
(0)	4	3)

the sequential presentation of treatments the participant's performance is evaluated, not *which* treatment is experienced.

Sequence or order effects will be experienced equally by all participants in counterbalanced situations because each treatment appears an equal number of times at each testing session. This consideration points to a major flaw in the use of randomized, incomplete counterbalancing: The treatments may not be presented an equal number of times at each testing session (see Table 6-5). Thus, sequence or order effects are not controlled in this situation.

Carryover effect The effects of one treatment persist or carry over and influence responses to the next treatment.

Carryover Effects When a **carryover effect** is present, the effects of one treatment continue to influence the participant's response to the next treatment. For example, let's assume that experiencing the green (G) warning light before the red (R) light always causes participants to *decrease* their reaction time by 2 seconds. Conversely, experiencing R before G causes participants to *increase* their reaction time by 2 seconds. Experiencing the flashing white (FW) warning light either before or after G has no effect on reaction time. However, experiencing R before FW increases reaction time by 3 seconds, and experiencing FW before R reduces reaction time by 3 seconds. In the R → G/G → R and R → FW/FW → R transitions, the previous treatment influences the participant's response to the subsequent treatment in a consistent and predictable manner. These effects are diagrammed in Table 6-8. Note that counterbalancing includes an

TABLE 6-8 Example of Carryover Effects in a Counterbalanced Experiment

Carryover effects occur when a specific preceding treatment influences the performance in a subsequent treatment. In this example, experiencing Treatment G prior to Treatment R results in a decrease of 2 (i.e., -2), whereas experiencing Treatment R prior to Treatment G results in an increase of 2 (i.e., +2). Experiencing Treatment G prior to FW or Treatment FW prior to G does not produce a unique effect. However, experiencing Treatment R prior to FW results in an increase of 3, whereas experiencing Treatment FW prior to R results in a decrease of 3.

	Sequence of Treatments		
	G	R	FW
Effect on Performance →	(0	-2	+3)
	G	FW	R
	(0	0	-3)
	R	G	FW
	(0	+2	0)
	R	FW	G
	(0	+3	0)
	FW	G	R
	(0	0	-2)
	FW	R	G
	(0	-3	+2)

equal number of each type of transition (e.g., R → G, G → R, R → FW, etc.). Thus the opposing carryover effects cancel each other.

Differential Carryover Although counterbalancing can control many things, it offers no protection against differential carryover. **Differential carryover** occurs when the response to one treatment depends on which treatment is experienced previously. Consider an experiment investigating the effects of reward magnitude on reading comprehension in second-grade children. Each child reads three similar passages. After each passage is completed, a series of questions is asked. Each correct answer is rewarded by a certain number of M&Ms (the IV). In the low-reward condition (A), children receive one M&M after each correct answer; three and five M&Ms are received in the medium-reward (B) and high-reward (C) treatments, respectively.

Differential carryover
The response to one treatment depends on which treatment was administered previously.

Although this experiment might be viewed as another six-sequence example of counterbalancing (see Table 6-4), the effects may not be symmetrical as in the carryover example we just considered. The participant who receives the A → B → C sequence may be motivated to do well at all testing sessions because the reward progressively increases from session to session. What about the student who receives the B → C → A sequence? In this case the reward rises from three M&Ms (Session 1) to five M&Ms (Session 2), but then is reduced to one M&M (Session 3). Will the decrease from five M&Ms to one M&M produce a unique effect—such as the student's refusing to participate—not seen in the other transitions? If so, differential carryover has occurred, and counterbalancing is not an effective control procedure. Some possible effects of differential carryover in the M&M study are shown in Table 6-9. As you can see, the

TABLE 6-9 Example of Differential Carryover in a Counterbalanced Experiment

Differential carryover occurs when performance depends on which specific sequence occurs. In the following example, experiencing Treatment A prior to Treatment B results in an increase of 6 (i.e., +6), whereas all other sequences result in an increase of 2 (i.e., +2).

Sequence of Treatments			
	A (1 M&M)	B (3 M&Ms)	C (5 M&Ms)
Effect on Performance →	(0	+6	+2)
A	C	B	
(0	+2	+2)	
B	A	C	
(0	+2	+2)	
B	C	A	
(0	+2	+2)	
C	A	B	
(0	+2	+6)	
C	B	A	
(0	+2	+2)	

drastic decrease in performance produced by the C to A (five M&Ms to one M&M) transition is not canceled out by a comparable increase resulting from the A-to-C transition.

The potential for differential carryover exists whenever counterbalancing is used. The experimenter must be sensitive to this possibility. A thorough review of the effects of the IV being used in your research can help sensitize you to the possibility of differential carryover. If this threat exists, it is advisable to employ a research procedure that does not involve presenting more than one treatment to each participant.

■ REVIEW SUMMARY

1. **Randomization** controls for extraneous variables by distributing them equally to all groups.
2. When experimenters use **elimination** as a control technique, they seek to remove completely the extraneous variable from the experiment.
3. **Constancy** controls an extraneous variable by creating a constant or uniform condition with regard to that variable.
4. **Balancing** achieves control of extraneous variables by ensuring that all groups receive the extraneous variables to the same extent.
5. **Counterbalancing** controls for sequence or order effects when participants receive more than one treatment. **Within-subject counterbalancing** involves the administration of more than one treatment sequence to each participant, whereas **within-group counterbalancing** involves the administration of a different treatment sequence to each participant.
6. The total number of treatment sequences can be determined by $n!$ When all sequences are administered, **complete counterbalancing** is being used. **Incomplete counterbalancing** involves the administration of fewer than the total number of possible sequences.
7. The random selection of treatment sequences, systematic rotation, or the Latin square approaches can be used when incomplete counterbalancing is implemented.
8. Counterbalancing can control for **sequence or order and carryover effects**. It cannot control for **differential carryover**, in which the response to one treatment depends on which treatment was experienced previously.

■ Check Your Progress

1. Matching
 1. randomization
 2. elimination
 3. constancy
 4. balancing
 5. counterbalancing
 - A. complete removal of the extraneous variable
 - B. extraneous variable is reduced to a single value
 - C. most widely used control procedure
 - D. used to control for order effects
 - E. extraneous variable is distributed equally to all groups

2. The most widely used control technique is
 - a. randomization
 - b. elimination
 - c. constancy
 - d. balancing
 - e. counterbalancing
3. Balancing is a logical extension of
 - a. constancy
 - b. randomization
 - c. counterbalancing
 - d. elimination
4. Distinguish between within-subject and within-group counterbalancing.
5. What can $n!$ be used for? Calculate the value of $4!$
6. _____ occurs when the response to one treatment depends on which treatment preceded it.
 - a. Carryover
 - b. Differential carryover
 - c. A sequence or order effect
 - d. Fault randomization
7. What is incomplete counterbalancing?

Key Terms

Variable, 100	Nuisance variable, 109	Complete counter-balancing, 119
Operational definition, 101	Randomization, 113	Incomplete counter-balancing, 119
Physiological IV, 102	Elimination, 114	Sequence or order effects, 121
Experience IV, 102	Constancy, 114	Carryover effect, 122
Stimulus or environmental IV, 102	Balancing, 115	Differential carryover, 123
Participant characteristics, 103	Counterbalancing, 117	
Confounding, 105	Within-subject counterbalancing, 117	
Valid, 108	Within-group counterbalancing, 117	
Reliable, 108		

Looking Ahead

In this chapter we have explored the nature of variables in general and have seen the importance of selecting appropriate IVs and DVs. The potentially damaging effects of nuisance variables and confounders have led to the development of procedures for their control. We continue our examination of the basics of experimentation in the next chapter. There we will discuss the selection of appropriate types and numbers of participants and the actual collection of research data.

Conducting a Good Experiment II: Final Considerations, Unanticipated Influences, and Cross-Cultural Issues

Participants

- Types of Participants
- Number of Participants

Apparatus

- IV Presentation • DV Recording

The Experimenter as an Extraneous Variable

- Experimenter Characteristics
- Experimenter Expectancies
- Controlling Experimenter Effects

Participant Perceptions as Extraneous Variables

- Demand Characteristics and Good Participants
- Response Bias • Controlling Participant Effects

The Interface Between Research and Culture

- Culture, Knowledge, and Truth
- The Effect of Culture on Research
- Methodology and Analysis Issues

In this chapter we continue our discussion of research basics by considering the type and number of participants to be used, the type of apparatus or testing equipment to be employed, whether the experimenter or participants may be extraneous variables, and the cross-cultural implications of research.

Participants

Both the type and the number of participants who take part in a research project are important considerations in the formulation of an experiment. We discuss these aspects in this section.

Types of Participants

A moment's reflection will indicate that there is a wealth of potential organisms that can serve as participants in a psychological experiment. For example, animal researchers might choose from bumblebees, flies, dolphins, chimpanzees, elephants, rats, and so on. Likewise, researchers dealing with humans might choose as their participants infants, young adults, the aged, the gifted, the handicapped, or the maladjusted, among many others. The range of participants for your research project may be overwhelming. Which one represents the best

choice? Considering three things can help you answer this question: *precedent*, *availability*, and the *nature of the problem*.

Precedent If your literature review indicates that researchers have worked with a particular type of participant successfully in prior research projects in your area of interest, you may want to consider using this type of participant. For example, when Willard S. Small (1901) conducted the first rat study in the United States, he began a **precedent**, an established pattern, that continues to this day. Likewise, the precedent for using college students (especially those enrolled in introductory psychology) has a venerable history. For example, research in the area of human memory and cognitive processes relies heavily on projects conducted with college students. How strong is this reliance? We selected a copy of the *Journal of Experimental Psychology: Learning, Memory and Cognition* (Vol. 24, No. 4, July 1998) at random and examined it. Of the 15 articles published in this issue, 12 reported results from experiments that used college students exclusively as participants; 3 reported using paid volunteers or participants who may have been college students. It is noteworthy that in many of the articles using college students, the participants were drawn from a subject pool, or they participated in order to receive partial course credit.

Precedent An established pattern.



Such precedents for the selection of participants have their advantages and disadvantages. What are they?

The fact that a particular type of participant has been used repeatedly in psychological research ensures that a body of knowledge exists about that type of participant. Researchers can take advantage of this wealth of knowledge as they plan their research. They can implement already validated procedures without expending hours of exploratory testing and designing new equipment. Being able to draw on this body of already proven techniques means the *likelihood of success* (see Chapter 2) is increased.

However, the continual use of one type or species of participant can limit the *generalizability* of the information that is gathered (see the discussion of external validity in Chapter 8). For example, although the study of self-esteem in college students may tell us about this trait in that group, it may not tell us much about self-esteem in the general population. Likewise, although the choice to use white rats in an experiment may be prompted by the extensive literature that already exists on this animal, additional rat studies may not provide any useful information about other species.

Availability The continued use of white rats and college students also stems from another source, their availability. White rats are relatively inexpensive, at least compared to many other animals, and they are easy to maintain. Likewise, college students, especially those students enrolled in introductory psychology classes, constitute an easily accessible population from which to draw participants. For example, in his study of male and female differences in altruism (unselfishly helping another person), Nicholas Schroeppel, a student at the

University of Nebraska in Lincoln, and his faculty advisor, Gustavo Carlo, administered questionnaires to college students (Schroeppel & Carlo, 2001). In addition to their availability, college students are inexpensive participants because researchers do not usually pay them for their participation. At some institutions participation in a research project may be a course requirement, thus ensuring the ready availability of research participants. Availability, of course, does not guarantee that the researcher has selected the best or most appropriate participants.

Clearly, availability of one type of participant may discourage the use of others. In turn, we have seen that developing a large body of knowledge about a particular type of participant can result in pressures to use that participant in future research projects. Obviously, the problem can easily become circular. The precedent for using one type of participant is established and leads to the development of an extensive literature concerning that type of participant, which further encourages the use of that type of participant.

Type of Research Project Often the nature of your research project will determine the type of participant you decide to use. For example, if you are interested in studying the visual ability of birds of prey, you are limited to studying birds such as eagles, vultures, hawks, and owls; ducks, geese, and songbirds are not predators. Likewise, if you want to study hallucinations and delusions, you have limited your choice of potential participants to humans who are able to communicate and who experience those phenomena.

Consider the research project conducted by Molly Claus (2000), a student at Nebraska Wesleyan University in Lincoln. She was interested in studying the relation between toy preferences and assertiveness. The nature of this research question dictated that she use children as participants; she studied preschoolers. Her research indicated that children who played with masculine toys (e.g., trucks and blocks) were more assertive than children who chose to play with feminine toys (e.g., a tea set).

Number of Participants

Once you have decided what type of participant to use in your research project, you must then determine how many participants you are going to test. In making this decision, there are numerous factors to take into account, including these:

1. **Finances.** How much will it cost to test each participant? Animals must be purchased and cared for. It may be necessary to pay humans for their participation. Does the person who actually conducts the experiment have to be paid? If so, this cost also must be considered; it will rise as you test additional participants.
2. **Time.** As you test additional participants, time requirements will increase, especially if you test participants individually.
3. **Availability.** The number of available participants may influence how many you choose to use in your experiment.

In addition to these more practical considerations, there is another factor that enters into our determination of the number of participants we will use. This factor is the amount of variability we expect to be present within each group. The less the *within-group* variability (i.e., the more homogeneous the participants), the fewer participants we will need. Conversely, the greater the *within-group* variability (i.e., the more heterogeneous the participants), the greater the number of participants we will need.



What is the reasoning behind these statements about variability and the number of participants to be tested in an experiment? In thinking about this question, you may want to review the material on nuisance variables (see Chapter 6).

When a nuisance variable, in this case the heterogeneous nature of the participants, is present, the scores within each group spread out considerably and the amount of group overlap is increased. The variability of the scores makes it more difficult to see absolute differences between the groups. One way to deemphasize these extreme scores is to test more participants. By increasing the number of scores (i.e., participants), we should increase the number of scores that cluster in the center of the distribution and, therefore, decrease the impact of the extreme scores. When nuisance variables are not present (i.e., the groups are more homogeneous), there are fewer extreme scores and the differences between the groups can be seen more clearly.

As you saw in Chapter 4, another way to create more homogeneous groups is to use stratified random sampling. By sampling a more specific type of participant, we remove extremes from our sample. We will have more to say about the number of observations and variability in the next chapter.

The number of participants tested is related to the *power* of our statistical test. **Power** is the likelihood (probability) that the statistical test will be significant (i.e., the experimental hypothesis is accepted when it is true). Generally speaking, the greater the number of participants, the higher the power of the statistical test; therefore, it is to your advantage to use as many participants as possible under the specific constraints of your research. Statistics books routinely discuss power and present formulas and tables that will aid you in determining the number of participants you will need to test in order to achieve significance.

Power The probability that a statistical test will be significant (i.e., the experimental hypothesis is accepted when it is true).

You should also use your literature review as a guideline concerning the number of participants you need to use in your experiment. If previous studies in your area have successfully used a certain number of participants, you can assume you will need to test a comparable number. For example, based on precedent, Kimberly Kiker (2001), a student at the University of Florida in Gainesville, chose to use 20 first-grade schoolchildren and 20 third-grade schoolchildren in her research on measurement errors in young children. If you are conducting a project in an area that has not received much research attention, however, you will want to test as many participants as possible given your financial, time, and availability constraints.

Apparatus

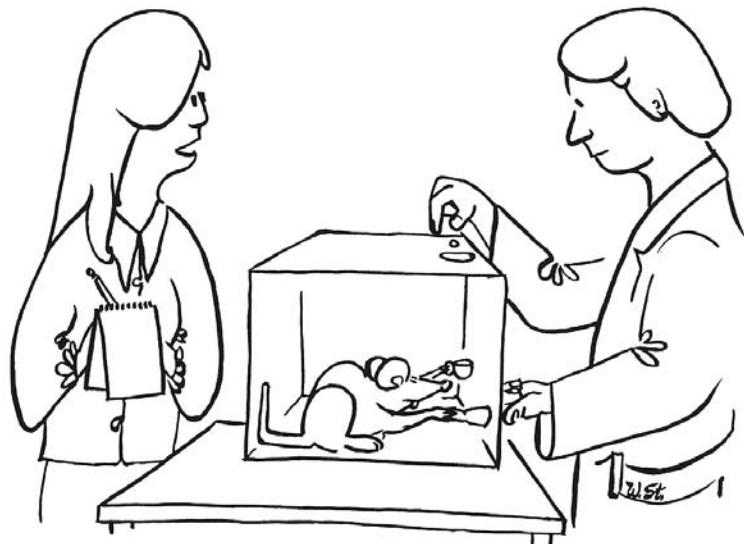
While you are deciding on the number and type of participants to use in your research, you also need to consider the type of apparatus, if any, you will be using. It is possible to use apparatus both for presenting the independent variable (IV) and for recording the dependent variable (DV).

IV Presentation

Often the nature of the IV will influence the type of apparatus you choose to use. For example, Michael Jones (2001), a student at Lincoln Memorial University in Harrogate, Tennessee,

was interested in studying the effects of noise and sex of participant on recall and spatial task (completing a block design) performance of children between 9 and 11 years old. He presented unstructured white noise (static) or a popular song to groups of children. His experiment dictated that he use audio equipment. He recorded the popular song on a continuous-loop tape that played the same song again and again. He presented the song and the white noise at exactly 74 decibels (dB), as measured by a digital sound meter. The results of this study indicated that participants in the white-noise condition performed better on both tasks than did the group of children who heard the popular song.

The possible ways in which you can present the IV are limited by the type of IV you are manipulating, by finances, and by your ingenuity. Clearly, the presentation of certain types of IVs requires specialized equipment. For example, the administration of a particular type of light will require a specialized projector. On the other hand, presenting food to a hungry rat that has learned to press a lever does not have to involve the purchase of an expensive Skinner box or a food-pellet dispenser. We have seen very effective handmade Skinner boxes and food-delivery systems. For example, a sturdy cardboard box works just fine in place of a commercial Skinner box. What about the lever? No problem. Anything that protrudes into the box that the rat can learn to press or touch will work fine. In fact, some students have simply drawn a circle on one wall and required that the rat touch it with its nose before food is delivered. A piece of plastic tubing can serve as the food-delivery system. Simply fasten the tubing to the outside of the box so that one end is at the top and the other end enters the box near the bottom and is situated over a small dish. Dropping a piece of food down the tube results in a reinforcer being delivered into the food dish in the box. Some creative thought at this stage of planning might save you substantial amounts of money.



First published in the APA Monitor. Reprinted by permission of Warren Street.

"Besides operant equipment, Professor, have the federal cutbacks hurt you?"

Although large research laboratories may require extensive funds and elaborate equipment, you can conduct meaningful research on a limited budget with little or no fancy equipment.

DV Recording

Although recording evidence in great detail at the scene of a crime may be a major concern to the detective, recording the DV is such a fundamental task that it sometimes is taken for granted or overlooked by the psychologist. However, there are problems and options to be addressed. For example, you do not want the researcher's presence to interfere with or modify the participant's behavior. Sometimes being inconspicuous may mean that the researcher will need to use a simple paper-and-pencil data sheet instead of elaborate, highly visible electronic equipment. In short, it is important for the observer to be inconspicuous while recording data.



Although technologically sophisticated equipment can be beneficial in conducting research, potential drawbacks are associated with its use. What are some of these drawbacks?

Whether presenting the IV or recording the DV, the experimenter should not become a slave to the equipment that is available. Just because you have access to a certain piece of equipment does not mean that you must use it. If a handheld stopwatch provides data equivalent to or better than what can be obtained with an elaborate computer setup, the experimenter should prefer the less elaborate equipment. If researchers begin to rely too heavily on their equipment, then the choice of research problems may start to be dictated by the equipment, not by the researchers' creativity. In such situations there is more concern for the IV that can be presented by the equipment than there is for the problem that is being investigated. Also, the presence of elaborate pieces of equipment assumes that the researcher has sufficient funds to provide appropriate maintenance for such equipment. Consider the problems you would face if your research depended on a particular piece of equipment that was broken and you had no funds to pay for its repair.

In some instances a good pair of eyes may be your best equipment. Consider the following exchange between Sherlock Holmes and Dr. Watson:

"What can you learn from this battered old felt?" Watson asks Holmes. "Here is my lens. You know the methods."

"I can see nothing."

"On the contrary, Watson, you can see everything." (Doyle, 1927, p. 246)

■ REVIEW SUMMARY

1. **Precedent**, availability, and the nature of the problem are factors that influence the choice of participants to use in psychological research.
2. The number of participants used in a research project will be determined by financial considerations, time constraints, and participant availability. Generally speaking, the larger the number of participants, the greater the **power**.

3. The use of homogeneous groups of participants allows the experimenter to test smaller groups, whereas heterogeneous groups dictate using more participants.
4. Automated equipment can be used for IV presentation and DV recording, but often quality data can be recorded with less sophisticated equipment.

■ Check Your Progress

1. Explain the relation between precedent and the type of participant used in a research project.
2. Why are white rats and college students the favorite participants in psychological research?
3. Group homogeneity is best associated with
 - a. testing a larger number of participants
 - b. testing a smaller number of participants
 - c. multiple-strata sampling
 - d. nonrandom sampling
4. One of the best guidelines for the number of participants to be used in a successful research project is
 - a. cost
 - b. availability
 - c. past research
 - d. type of IV presentation
5. You would like to assess the effects of trauma in people's lives, but you have ethical problems with causing traumatic events to occur to people. You have a problem with
 - a. IV presentation
 - b. IV recording
 - c. DV presentation
 - d. DV recording
6. Describe the concern the experimenter must be aware of when using automated equipment to present the IV or the DV.

Although you may believe that you are ready to move forward with your research project, we still have some preliminary issues to consider. Just as the good detective is careful, precise, and thorough, researchers cannot exercise too much care in developing their experiments. In the following sections we highlight two potential extraneous variables that often go overlooked: the experimenter and the participant. As the science of psychology matures, researchers pay increasing attention to these factors.

The Experimenter as an Extraneous Variable

Just as the characteristics of the detective can influence the responses of the suspects who are questioned, several aspects of the experimenter can influence the responses of the participant (Rosenthal, 1976). First we explore experimenter characteristics; then we will consider experimenter expectancies.

Experimenter Characteristics

Both physiological and psychological attributes of the experimenter can influence the outcome of your research. Physiological characteristics include such variables as age, sex, and race. Research has shown that each of these variables can have an influence on participants' responses. For example, Robert Rosenthal (1977) has shown that male experimenters are more friendly to their participants than female experimenters.



If all participants in an experiment are tested by the same experimenter, then constancy (see Chapter 6) is established, and this extraneous variable is controlled. If such control is exercised, why are experimenter characteristics, such as age, sex, and race, of potential concern?

Even though constancy is achieved by having one experimenter conduct the research project, it is possible that the experimenter will influence the participants in a unique manner. Perhaps the friendliness of a male experimenter will encourage all participants to perform at a very high level, thus making the differences between the treatment groups less evident. Hence, problems can arise when one attempts to compare the results of similar research projects conducted by different experimenters. If these projects yield different results, you cannot be sure whether the differences are attributable to differences in the IV or to effects created by the different experimenters.

The psychological attributes of the experimenter that can influence the results of an experiment include personality characteristics such as hostility, anxiety, and introversion or extraversion. An experiment conducted by an experimenter who is highly anxious is likely to yield different results than an experiment conducted by a confident, self-assured experimenter. The same is true for other personality characteristics.

Experimenter Expectancies

In addition to physiological and psychological characteristics, the experimenter's expectations concerning the participants' behavior can, and do, affect performance. The experimenter's expectations cause him or her to behave toward the participants in such a manner that the expected response is, indeed, more likely shown. The experimenter is literally a cause of the desired or expected response.



Are experimenter expectancies best categorized as extraneous variables that can confound the experiment, or are they simply nuisance variables that obscure the effects of the IV?

If experimenter expectancy is operating in your experiment, you cannot tell whether your results are due to the influence of the IV or experimenter expectancy; hence, they are best labeled as confounders.

Such effects have been demonstrated in both human and animal experiments. One of the most widely cited studies of the effects of human experimenter expectancy involved the IQ scores of grade-school children (Rosenthal & Jacobson, 1968). At the start of the school year all children in the classes that were studied took an IQ test. Then, several children in each class were randomly selected, and their respective teachers were told that these children were "intellectual bloomers." Several months later, when all the children retook the IQ test, the IQ scores of the intellectual bloomers had increased more than those of the other children. Because the intellectual bloomers were randomly selected, it is doubtful that they were intellectually superior. However, they were perceived in this manner and were treated differently by the teachers. In turn, these students responded in accordance with the teachers' expectations.

Experimenter expectancy is not limited to studies of humans; even rats will perform in accordance with what the experimenter anticipates. Rosenthal and Fode (1963) told half the students in a class that the rats they were going to train were "maze bright"; the remainder of the students were told that their rats were "maze dull." In actuality, there were no differences among the rats at the start of the project. The results, however, were consistent with the students' expectations: The "smart" rats learned the maze better and more quickly than did the "dumb" rats. Because the rats did not differ at the start of training, this study clearly highlights the strong effects that experimenter expectancies can have on participants' performance. Because Rosenthal and his colleagues were among the first systematically to study experimenter expectations, the results of such expectations are often called **Rosenthal effects**.

Rosenthal effect The result when an experimenter's preconceived idea of appropriate responding influences the treatment of participants and their behavior.

Controlling Experimenter Effects

Physiological and Psychological Effects The reason for experimenters traditionally paying little attention to these variables may be clear by now: They are difficult to control. For example, to achieve constancy, *all* these characteristics would have to be measured in *all* potential experimenters and then the choice of

experimenters would be determined by the level of each factor that was desired—a difficult, if not impossible, task. Likewise, we saw in Chapter 6 that balancing can be used to avoid confounding caused by the sex of the experimenter. Although this control technique equates the groups for experimenter sex, it does not simultaneously control for other physiological and psychological characteristics. At present, the most common procedures for controlling general experimenter characteristics are to (a) use standardized methods, (b) use careful training to a set standard when the experimenter administers procedures, and (c) standardize appearance, attitude, and so forth as much as possible. Replicating your research also provides a good check on possible experimenter effects; if you replicate the findings with a different experimenter, then experimenter effects are less likely to be a factor. A thorough literature review will help make you aware of any relevant experimenter variables in your area of research interest.

Experimenter Expectancies Several things can be done to reduce, or possibly eliminate, experimenter expectancies. First, the instructions that the experimenter gives to the participants should be carefully prepared so their manner of presentation will not influence the participants' responses. Likewise, any instructions concerning scoring the participants' responses should be as objective and concrete as possible and established before beginning

the experiment. If these instructions are subjective, then there is room for experimenter expectancies to dictate how they will be scored.

A second method for controlling experimenter expectancies involves the use of instrumentation and automation. For example, instructions to participants may be tape-recorded prior to experimental testing in order to reduce any influences the experimenter might have. In many instances, potential influences of the experimenter are eliminated through the use of printed instructions or computer displays. Also, automated equipment can ensure the accurate recording and storage of response data. In some instances, the participants' responses are entered directly at a computer terminal and thus are stored and ready for analysis at any time.

A third method for minimizing experimenter expectancies is to conduct a single-blind experiment. The **single-blind experiment** keeps the experimenter in the dark regarding which participants receive which treatment(s). (As we will see, this procedure can be used to control participant effects as well.) For example, suppose you are conducting an experiment testing the effects of different descriptions of an individual's degree of altruism on the amount of warmth that individual is perceived to possess. Quite likely the different written descriptions will be printed in such a way that they all have the same appearance (constancy). If these different descriptions have the same demographic cover sheet, then all materials will appear to be identical. If another experimenter is in charge of determining which participants read which descriptions, and arranging the order in which to distribute the testing materials to the participants, then the experimenter who actually conducts the research sessions will not know which descriptions are being read by which participants at any session, and therefore experimenter expectancies cannot influence the participants' responses. The single-blind procedure can also be used effectively when the IV consists of some type of chemical compound, whether it be in tablet or injection form. If all tablets or injections have the same physical appearance, the experimenter will be unaware of which participants are receiving the active compound and which participants are receiving a placebo. The experimenter truly does not know what treatment condition is being administered in single-blind experiments of this nature.

Single-blind experiment
An experiment in which the experimenter (or participants) is unaware of the treatment the participants are receiving.

Participant Perceptions as Extraneous Variables

Just as the experimenter can unintentionally influence the results of an experiment, so can the participants. As you will see, numerous aspects of the participants' perception of the research project can operate as extraneous variables and nuisance variables.

Demand Characteristics and Good Participants

If you have ever served as a participant in a psychological experiment, you know that most participants believe that they are supposed to behave in a certain manner. As we have seen, the cues used by participants to guide their behavior may come from the experimenter; they may also be part of the experimental context or IV manipulation. When participants use these cues to determine what the experimenter's hypothesis is and how they are supposed to act, the cues are referred to as the **demand characteristics** of the experiment (Orne, 1962). In short, participants in psychological research may attempt

Demand characteristics
Features of the experiment that inadvertently lead participants to respond in a particular manner.



Reprinted from The Chronicle of Higher Education. By permission of Mischa Richter and Harold Bakken.

"Haven't we met in a previous experiment?"

It is important that participants not communicate about the nature of a psychological experiment.

to figure out how they are supposed to respond and then behave in this manner. The desire to cooperate and act as the participants believe the experimenter wants them to is called the **good participant effect** (Rosenthal & Rosnow, 1991).

Allison Dickson, James Morris, and Keri Cass, students at Southwestern University in Georgetown, Texas, and their faculty advisor, Traci Giuliano, investigated race, stereotypes, and college students' perceptions of rap and country music performers. The "participants made judgments about either a Black or a White musician who performed either rap or country music" (Dickson, Giuliano,

Morris, & Cass, 2001, p. 175). Rather than tell the participants that they were studying the effects of race and stereotypes, the researchers simply "asked them to contribute to an investigation exploring 'people's perceptions of music'" (Dickson et al., 2001, p. 177). Had they divulged the exact nature of their hypothesis, then the demand characteristics would have been quite strong, and the participants might have tried to decide which group they were in and how they were expected to respond. Dickson et al. (2001) found that the participants rated the black rap performer more favorably than the black country performer. The participants also rated the white country performer more favorably than the white rap performer.



Although experimenters have an ethical responsibility (see Chapter 2) to inform the participants about the general nature of the experiment, we have just seen that they usually do not want to reveal the exact nature of their experimental hypothesis. To do so can introduce strong demand characteristics that could influence the participants' behavior. Do demand characteristics operate as extraneous variables or nuisance variables?

Depending on how the demand characteristics are perceived, they can operate either as an extraneous or nuisance variable. If it is very clear to the participants which group they are in and how that group is expected to act, then the demand characteristics are functioning as an extraneous variable, and the experiment is confounded. When the experiment is completed, the experimenter will not be able to tell whether the differences between the groups were due to the effects of the IV or the demand characteristics. If, however, the participants perceive the demand characteristics *but* are not sure which group they are in, then the demand characteristics may function as a nuisance variable. In this situation we would expect the demand characteristics to produce both increases and decreases in responding within all groups. Thus, the scores of *all* groups would spread out more.

Response Bias

Several factors can produce a response bias on the part of research participants. Here we examine such influences as yea-saying and response set.

Yea-Saying You probably know individuals who seem to agree with everything, even if agreeing sometimes means that they contradict themselves. We may never know whether these individuals agree because they truly believe what they are saying or because they are simply making a socially desirable response at the moment. Clearly, these individuals, known as **yea-sayers**, who say yes to all questions, pose a threat to psychological research. (Individuals who typically respond negatively to all questions are known as **nay-sayers**.) Contradicting one's own statements by answering yes (or no) to all items on a psychological inventory or test seriously threatens the validity of that participant's score.

Yea-sayers Participants who tend to answer yes to all questions.

Response Set Sometimes the experimental context or situation in which the research is being conducted can cause participants to respond in a certain manner or have a **response set**. The effects of response set can be likened to going for a job interview: You take your cues from the interviewer and the surroundings. In some cases, you will need to be highly professional; in other interview situations, you can be a bit more relaxed.

Nay-sayers Participants who tend to answer no to all questions.

Consider the two following descriptions of an experimenter and the testing room in which the research is being conducted. In the first instance, the experimenter is wearing a tie and a white laboratory coat. The experiment is being conducted in a nicely carpeted room that has pleasing furnishings, several bookcases, and attractive plants; it looks more like an office than a research room. The second experimenter is dressed in jeans, a sweatshirt, and tennis shoes. In this case the research is being conducted in a classroom in a less-than-well-kept building. Have you already developed a response set for each of these situations? Will you be more formal and perhaps give more in-depth or scientific answers in the first situation? Even though the second situation may help put you at ease, does it seem less than scientific? Notice that our descriptions of these two situations did not make reference to the physiological or psychological characteristics of the experimenters or to the type of experiment being conducted. Hence, we are dealing with an effect that occurs in addition to experimenter and demand effects.

Response set The result when an experimental context or testing situation influences the participants' responses.

Likewise, the nature of the questions themselves may create a response set. For example, how questions are worded or their placement in the sequence of items may prompt a certain type of response; it may seem that a socially desirable answer is being called for. Also, it may seem that a particular alternative is called for. In such cases a response bias is developed. Clearly, response set can be a major influence on the participant's response.

Controlling Participant Effects

As we have seen, there are aspects of the participant that can affect the results of our research. Although such factors are rather difficult to control, researchers have developed several techniques.

Demand Characteristics You will recall that one technique used to control for experimenter expectancies is to keep the experimenter in the dark by conducting a single-blind experiment. This same approach can be applied to the control of demand characteristics, only this time the participants will be unaware of such features as the experimental hypothesis, the true nature of the experiment, or which group they happen to be in.

It takes only a moment's thought to reach the conclusion that these two approaches can be combined; we can conduct an experiment in which both the experimenter and the participants are unaware of which treatment is being administered to which participants. Such experiments are known as **double-blind experiments**.

Regardless of whether a single- or double-blind experiment is conducted, it is likely that the participants will attempt to guess what the purpose of the experiment is and how they should respond. It is difficult to conceal the fact that they are participating in an experiment, and the information provided them prior to signing the informed consent document (see Chapter 3) may give the participants a general idea concerning the nature of the experiment.

Double-blind experiment

An experiment in which both the experimenter and the participants are unaware of which treatment the participants are receiving.



Let's assume that you are conducting a single- or double-blind experiment and you are leaving your participants to their own devices to guess what the experiment concerns. In this case you may be introducing an unwanted variable into your research project. Is this factor an extraneous variable or a nuisance variable? What effect(s) might it have?

It is almost certain that all participants will not correctly guess the true nature of the experiment and which group they are in. Those participants who make correct guesses may show improved performance; those participants who make incorrect guesses may have inferior performance. If the ability to guess correctly the nature of the experiment and which group one is in is comparable in all groups, then the scores within *all* groups will spread out. You have introduced a *nuisance variable* that will make it more difficult to see the differences that develop between the groups. Is there any way to avoid this problem? The answer is yes.

Another technique that can be used to control for demand characteristics is to give all participants incorrect information concerning the nature of the experiment. In short, the experimenter purposely deceives *all* the participants, thus disguising the true nature of the experiment and keeping the participants from guessing how they are supposed to respond.

Although this procedure can be effective, it suffers from two drawbacks. First, we have already seen that the use of deception raises ethical problems with regard to the conduct of research. If deception is used, a good Institutional Review Board (IRB; see Chapter 2) will be careful to make sure it is a justifiable and necessary procedure. The second problem is that the information used to deceive the participants may result in erroneous guesses about the nature of the experiment; the participants are then responding to the demand characteristics created by the deception. Clearly, demand characteristics may be very difficult to control.

Yea-Saying The most typical control for yea-saying (and nay-saying) is to rewrite some of the items so that a negative response represents agreement (control for yea-saying) or a positive response represents disagreement (control for nay-saying). After some of the items have been rewritten, the experimenter needs to make a decision concerning the order for their presentation. All the rewritten items should not be presented as a group. One presentation strategy is to randomize the complete list, thereby presenting the original and rewritten items in an undetermined sequence. This approach works quite well with longer lists. If the list is smaller, within-subject counterbalancing can be used. Table 7-1 illustrates these two presentation styles and the use of within-subject counterbalancing.

Response Set The best safeguard against response set is to review all questions that are asked or items to be completed to determine whether a socially desired response is implied in any manner. The answer given or response made should reflect the participant's own feelings, attitudes, or motives rather than an attempt to appear intelligent or well-adjusted or otherwise "normal." Checking for response set offers excellent opportunities for pilot testing and interviewing of participants to determine whether the questions or behavioral tasks create a particular outlook. Additionally, you should carefully examine the nature of the experimental situation or context to avoid the presence of undesired cues.

TABLE 7-1 Controlling for Yea-Saying

The following yes–no items are based on the Type A scale developed by Friedman and Rosenman (1974). In Part A a "yes" answer is associated with the Type A personality. Yea-sayers, even though they are not Type A persons, would likely be assigned to this category with such questions. In Part B half of the items are rewritten so that a "no" answer is associated with Type A characteristics. Within-subject counterbalancing is also used with the items in Part B.

- A. "Yes" answers are associated with Type A characteristics.
 1. Do you play games like Monopoly to win?
 2. Do you eat, speak, and walk rapidly?
 3. Do you constantly schedule more activities than time allows?
 4. Do you feel a sense of guilt if you try to relax?
- B. "Yes" answers (Items 1 and 4) and "no" answers (Items 2 and 3) are associated with Type A characteristics. The 1(yes)–2(no)–3(no)–4(yes) sequence illustrates within-subject counterbalancing.
 1. Do you play games like Monopoly to win?
 2. Do you eat, speak, and walk slowly?
 3. Do you constantly schedule just enough activities to fill the time available?
 4. Do you feel a sense of guilt if you try to relax?

■ REVIEW SUMMARY

1. Experimenter characteristics can affect the results of an experiment. Physiological experimenter characteristics include such aspects as age, sex, and race. Psychological experimenter attributes include such personality characteristics as hostility, anxiety, and introversion or extraversion.
2. Experimenter expectancies can produce behaviors in the experimenter that influence participants to make the desired response. Such experimenter expectancy effects are often called Rosenthal effects.
3. Because of their potential abundance, experimenter characteristics are difficult to control.
4. Experimenter expectancies can be controlled through the use of objective instructions and response-scoring procedures, as well as instrumentation and automation. A **single-blind experiment**, in which the experimenter does not know which participants are receiving which treatments, also can be used to control experimenter expectancies.
5. **Demand characteristics** refer to those aspects of the experiment that may provide the participants with cues concerning the experimenter's hypothesis and how they are supposed to act. Demand characteristics can be controlled through the use of single-blind and **double-blind experiments** (in which both the experimenter and the participants do not know which treatment the participants are to receive).
6. The desire to cooperate and act in accordance with the experimenter's expectation is called the **good participant effect**.
7. Response bias is caused by several factors. **Yea-saying** refers to the tendency to answer yes to all questions; **nay-saying** refers to the tendency to answer no to all questions. Yea- and nay-saying are controlled by writing some items in such a way that a negative response represents agreement (control for yea-saying) and a positive response represents disagreement (control for nay-saying).
8. When the experimental situation or context prompts a certain response, a **response set** has been created. The best safeguards against a response set are careful scrutiny of the experimental situation, thorough review of all questions, pilot testing, and interviewing the participants.

■ Check Your Progress

1. Explain how the experimenter can be an extraneous variable.
2. Matching
 1. age, sex, race
 2. hostility or anxiety
 3. Rosenthal effects
 4. single-blind experiment
 5. double-blind experiment
 - A. psychological experimenter effects
 - B. experimenter expectancies
 - C. physiological experimenter effects
 - D. control for demand characteristics and experimenter expectancies
 - E. control for experimenter expectancies

3. An experiment in which the researcher doesn't know which treatment the participants are receiving
 - a. is a blind experiment
 - b. is a confounded experiment
 - c. cannot be replicated
 - d. is unable to employ randomization
4. Explain how instrumentation and automation can control experimenter expectancies.
5. Demand characteristics refer to
 - a. demands the experimenter places on the participants
 - b. cues that tell the participants how to act
 - c. IRB requirements for conducting research
 - d. the need to publish research in order to receive tenure and promotion
6. Which of the following is an example of response bias?
 - a. demand characteristics
 - b. single-blind experiment
 - c. nay-saying
 - d. balancing

The Interface Between Research and Culture

The recent and continuing explosion in international information technology, coupled with the increased availability of air travel, highlights the diverse, multicultural nature of our planet. For example, it is common to see live reports on the television evening news from countries on the other side of the world. In our own country the “information superhighway” instantly links diverse peoples in the farthest reaches of the United States.

As we complete the material on the basics of experimentation, it is important to keep such cultural differences in mind; they may influence how research is conceptualized, conducted, analyzed, and interpreted. To this end, cross-cultural psychology has grown dramatically in recent years. The goal of **cross-cultural psychology** is to determine whether research results and psychological phenomena are universal (found in individuals from different cultures) or specific to the culture in which they were reported. Before we examine cross-cultural research, we need a definition of culture.

Cross-cultural psychology
A branch of psychology whose goal is to determine the universality of research results.



Rather than simply giving you a definition of culture, we would like you to spend a few minutes on this topic. Before reading further, decide which aspects or features should be included in a definition of culture.

You probably began thinking about culture with the word *similar*. After all, it's the similarities that help define our culture and distinguish it from other cultures. If you then started listing the important aspects that serve to distinguish cultures, you're on the right track. Among the important features that differentiate one culture from another are attitudes, values, and behaviors. Moreover, these culture-defining attitudes, values, and behaviors must

be long-lasting or enduring. This enduring quality indicates that these attitudes, values, and behaviors are communicated or transmitted from one generation to the next. Putting these considerations together, we can tentatively define **culture** as the lasting values, attitudes, and behaviors shared by a group of people and transmitted to subsequent generations. It is important to note that our definition does not imply that race and nationality are synonymous with culture. Individuals can be of the same race or nationality and not share the same culture. Carrying this thought one step further, we can see that several cultures may exist within the same country or even in the same large city. Now, let's see how culture is related to what we consider to be truth.

Culture, Knowledge, and Truth

Etic A finding that is the same in different cultures.

Emic A culture-specific finding.

A finding that occurs across cultures is called an **etic**. You can think of an etic as a universal truth or principle. The finding that reinforcement increases the probability of the response it follows appears to be an etic; people of all cultures respond to reinforcement similarly. In contrast, an **emic** is a finding that is linked to a specific culture. The value placed on independence and individuality is an emic; it varies from culture to culture. Some cultures (e.g., the United States) are individualistic and place a premium on individual accomplishments. Other cultures (e.g., China) are collectivist and stress contributions to the welfare of the group. Emics represent truths that are relative to different cultures, whereas etics represent absolute truths. Given the great diversity of cultures, it should not surprise you to find that the number of emics is considerably greater than the number of etics.



At this point you may be wondering how this discussion relates to a course in research methods or experimental psychology. Consider this issue before reading further.

Ethnocentric Other cultures are viewed as an extension of one's own culture.

If the goal of your research project is to discover the effects of IV manipulations only within a certain culture, then this discussion may have little relevance. However, few researchers purposely set their research sights on a single culture (see also Chapter 8). In fact, the question of culture may never enter the researcher's mind as a study is taking shape. Similarly, when the data are analyzed and conclusions reached, cultural considerations may not be addressed. The result is a project that one *assumes* is not culture dependent. In short, researchers often become **ethnocentric**—they view other cultures as an extension of their own. Hence, they interpret research results in accord with the values, attitudes, and behaviors that define their own culture and assume that these findings are applicable in other cultures as well. For example, many researchers probably consider the *fundamental attribution error* (Ross, 1977), in which actions are attributed to the individual even in the face of compelling situational cues, to be a universal finding. Although the fundamental attribution error is shown in individualistic, Western societies (Gilbert & Malone, 1995), situational cues are taken into account in collectivist societies, and this effect is appreciably smaller (Miller, 1984). We will have more to say about the generalizability of research findings in Chapter 8 when we consider the external validity of our research.

The Effect of Culture on Research

If you step back from your culture and try to put research in a more international perspective, it is clear that culture influences all aspects of the research process. We will consider cultural effects on the choice of the research problem, the nature of the experimental hypothesis, and the selection of the IV and the recording of the DV. To solve their respective problems, the detective and the research psychologist need a broad base that includes cultural information.

Choice of the Research Problem In some cases there may be no doubt that the choice of your research problem is culture dependent. For example, let's assume you are interested in studying the nature of crowd interactions at a rock concert. Whereas this topic may represent a meaningful project in the United States, it has much less relevance to a psychologist conducting research in the bush country of Australia. In this example, culture clearly dictates the nature of the research project; some problems are important in one culture but not in another. Likewise, whereas the study of individual achievement motivation is a topic of considerable interest and importance in an individualistic society like the United States, it would be a less important research topic in a collectivist society like China (Yang, 1982).

Nature of the Experimental Hypothesis When you have selected a problem that is relevant beyond your own culture, you must deal with the experimental hypothesis. For example, even though the study of factors that determine one's personal space is relevant in a number of cultures, the creation of an experimental hypothesis that applies to all cultures will be most difficult. In some cultures, very little personal space is the norm, whereas considerable personal space is expected in other cultures. For example, Italians typically interact within a smaller personal space (i.e., they prefer closer proximity) than either Germans or Americans (Shuter, 1977). Such cultural differences may lead to very different hypotheses.

Selection of the IV and Recording of the DV Culture also can influence the selection of the IV and the DV. In technologically advanced cultures such as those of the United States, Japan, or Great Britain, IV presentation may be accomplished via a computer. Likewise, DV measurement and recording may be conducted by computer. Because such technology is not available in all cultures, the choice of the IV and the procedure for recording the DV may differ considerably. For example, handheld stopwatches, not digital electronic timers, may be used to record the time required to complete an experimental task. Similarly, the participants may read the stimulus items in booklet form rather than having them presented at set intervals on a video screen by a computer. In fact, stimuli that have high (or low) meaning in one culture may not have the same meaning in another culture.

Methodology and Analysis Issues

In either conducting or evaluating cross-cultural research, a number of methodological issues will necessitate careful and thorough consideration. Among these issues are the participants and sampling procedures used, the survey or questionnaire employed, and the effects of cultural response set on data analysis.

Participants and Sampling Procedures The basic question here is whether the sample of participants is representative of the culture from which they were drawn. Do sophomore college students represent the culture of the United States? What steps have been taken to ensure that the sample is representative of the culture in question? For example, extreme differences may exist between samples drawn from large urban centers and those drawn from rural areas.

Assuming you can satisfy the requirement that a sample is representative of its culture, you are likely to be faced with an equally difficult task: being able to ensure that samples from two or more cultures are equivalent before conducting the research. We have stressed, and will continue to stress, the importance of establishing group equivalence before an IV is administered. Only when group equivalence is demonstrated can we have confidence in saying that our IV has had an effect in producing any differences we observe.



Let's assume you are reading research reports from three different cultures. All three investigations used freshman-level college students as participants. Can you assume that the requirements of group equivalence among these studies has been satisfied through the use of this common type of participant?

Even though the same *type* of participant—freshman college students—was used, you cannot assume the groups were equivalent before the research was conducted. Before you can make this assumption, you must be able to demonstrate that the same type of student attends college in all three cultures and that the “freshman” designation is equivalent across all three cultures. Perhaps the collegiate experience in one culture is quite different from the collegiate experience in another culture and therefore attracts students who differ from those in the culture with which it is being compared. For example, economics may dictate that only the wealthy attend college in some cultures. In short, assessing group equivalence will not be an easy task for cross-cultural research.

Type of Survey or Questionnaire Used Although an existing survey or questionnaire may work in a few instances, the researcher probably will not be able to use it for research in a different culture. The chances are good that the questionnaire or survey will have to be translated into a different language. Assuming the translation has been completed, how do you know that it is accurate and the correct meanings have been retained? The same word can have different meanings in different cultures. One technique for evaluating the accuracy of the translation is to have a *back translation* done. This procedure involves having another person translate the translated questionnaire back into its original language. If the back translation and the original version match, then the original translation was successful. The importance of the back translation is clearly reflected by Sou and Irving in their comments on the limitations of their research on student attitudes toward mental health in the United States and Macao (see Chapter 2). These researchers stated that “the survey items created for this study did not receive a backward translation to ensure reliability between the Chinese and English versions; therefore, group differences may be due to inherent differences in the meaning of the survey items after translation” (Sou & Irving, 2002, p. 21).

Aside from these problems, which are faced in translating the instrument into another language, there is the very real problem of whether the other cultures to be tested value the concept that is to be measured or evaluated. If they do not, then the survey or questionnaire is worthless.

Even if you determine that the concept or trait in question is valued in other cultures, there remains the problem of making sure that the *specific items* are equivalent when the survey or questionnaire is prepared for use in other cultures. Just translating the items into another language may not be sufficient. For example, a question about riding the subway, which is appropriate for an industrialized society, may have no meaning in a less-developed culture or even in some parts of the industrialized society. Clearly, the same comment can be made for questions dealing with customs, values, and beliefs.

Cultural Response Set Earlier in this chapter you learned that research participants may begin an experiment with a preexisting response set; some participants may be yea-sayers and others are nay-sayers. We have the same general concern, only on a larger scale, when conducting cross-cultural research.

In this instance it is the response of the entire culture, not individual participants, that concerns us. A **cultural response set**, or tendency of a particular culture to respond in a certain manner, may be operative. How often have individuals in another culture answered questions on a Likert-type rating scale like those commonly used in the United States? What reaction(s) will they have to being tested by such a scale?

The fact that you use such scales effectively in your research does not mean participants in other cultures will find them easy to understand and answer. The same comments can be made about any survey or questionnaire. Both the type of questionnaire (Likert-type scale, true-false, multiple choice, etc.) and the nature of the items themselves may intensify an already existing cultural response set.

How do you know whether a cultural response set is present? If differences exist among the groups tested in various cultures, a cultural response set may be operating.

Cultural response set

The tendency of a particular culture to respond in a certain manner.



The presence of a cultural response set is one possible cause for the differences among groups from various cultures. What other factor might cause such differences? What problem is created if you cannot distinguish between these two causes?

If you indicated that the influence of a manipulated IV or differences in the specific trait being measured could also be responsible for the differences among groups, you are absolutely correct. If you then indicated that an extraneous variable is present and the research would be confounded (see Chapter 6) if the researcher could not distinguish whether the scores on a questionnaire were caused by the trait or IV or the cultural response set, then you are correct again. Remember that our research results are worthless when a confounder is present. Hence, it is vitally important that cultural response set be accounted for whenever you are conducting or evaluating cross-cultural research.

The purpose of this section was not to teach all the fine points of cross-cultural research; it would take an entire book to accomplish that goal. Rather, we wanted to make you aware of these issues and problems before we began our discussions of internal and external validity, statistics, and research designs. Being acquainted with the issues involved in cross-cultural research will make you a better consumer of psychological research, regardless of where it is conducted.

■ REVIEW SUMMARY

1. The goal of **cross-cultural psychology** is to determine whether research findings are culture specific (i.e., **emics**) or universal (i.e., **etics**).
2. **Culture** can influence the choice of a research problem, the nature of the experimental hypothesis, and the selection of the IV and the DV.

3. In conducting or comparing cross-cultural research, the cultural representativeness of the participants and the sampling procedures used to acquire the participants must be carefully evaluated.
4. The appropriateness of a survey or questionnaire for cross-cultural research must be evaluated. When the general trait or concept has been accepted, the specific items must be examined and deemed acceptable for cross-cultural use.
5. The presence of a **cultural response set** must be considered in conducting cross-cultural research.

Check Your Progress

1. The goal of cross-cultural psychology is best described as
 - a. determining whether psychological findings are universal
 - b. testing a variety of participants
 - c. using a variety of test equipment
 - d. conducting tests in a variety of locations
2. A finding linked to a specific culture best describes

a. ethnocentrism	c. an emic
b. an ertic	d. an etic
3. Why is the goal of cross-cultural research incompatible with ethnocentrism?
4. In what ways can culture affect the conduct of psychological research?
5. The tendency of a culture to behave in a certain way best describes

a. an etic	c. ethnocentrism
b. a cultural response set	d. the just world stereotype

Key Terms

Precedent, 127	Yea-sayers, 137	Etic, 142
Power, 129	Nay-sayers, 137	Emic, 142
Rosenthal effect, 134	Response set, 137	Ethnocentric, 142
Single-blind experiment, 135	Double-blind experiment, 138	Cultural response
Demand characteristics, 135	Cross-cultural psychology, 141	set, 145
Good participant effect, 136	Culture, 142	

Looking Ahead

So far, our view of research in psychology has been rather general. At this point in the book we are on the verge of discussing specific research designs. (Our consideration of research designs begins in Chapter 10.) Before moving to this topic we will consider the internal and external validity of our research (Chapter 8) and the use of statistics in psychological research (Chapter 9).

Internal and External Validity

CHAPTER
8

Internal Validity: Evaluating Your Experiment From the Inside

- Threats to Internal Validity
- Protecting Internal Validity

- Threats to External Validity (Based on Our Participants)
- The Devil's Advocate: Is External Validity Always Necessary?

External Validity: Generalizing Your Experiment to the Outside

- Threats to External Validity (Based on Methods)

Internal Validity: Evaluating Your Experiment from the Inside

The concept of **internal validity** revolves around the question of whether your independent variable (IV) actually created any change that you can observe in your dependent variable (DV). Researchers take many precautions to increase internal validity as they design and set up experiments. If you use adequate control techniques, your experiment should be free from confounding and you can indeed conclude that your IV caused the change in your DV.

Internal validity A type of evaluation of your experiment; it asks whether your IV is the only possible explanation of the results shown for your DV.

Threats to Internal Validity

In this section we will alert you to several categories of factors that can cause problems with the internal validity of an experiment. These factors are details that you should attend to when planning and evaluating your experiment and when scrutinizing research done by other investigators. Sherlock Holmes stressed the importance of details to his assistant, Dr. Watson: "Never trust to general impression, my boy, but concentrate yourself upon details" (Doyle, 1927, p. 196). The following examples are drawn largely from the work of Donald T. Campbell (1957) and his associates (Campbell & Stanley, 1966; Cook & Campbell, 1979).

History We must beware of the possible extraneous variable of **history** when we test or measure experimental participants more than once. We might administer the IV between two DV measurements (pretest and posttest), or we may be interested in the time course of the

History A threat to internal validity; refers to events that occur between the DV measurements in a repeated-measures design.

IV effect—how participants react to the IV over time. In this case history refers to any significant event, other than the IV, that occurs between DV measurements. Campbell and Stanley (1966) cited the example of a 1940 experiment concerning informational propaganda. Researchers measured attitudes, and then students read some Nazi propaganda. Before the posttest was given, however, France fell to the Nazis. The subsequent attitude measurement was probably much more influenced by that historical event than by the propaganda. Obviously, this is an extreme example; historical events of this impact are rare. Campbell and Stanley also used the term to refer to less significant events that might occur during a lengthy experimental session. Noise or laughter or other similar events, which could occur during a session, might distract participants and cause them to react differently from participants who did not experience the same events. The distraction would then serve as an extraneous variable, and the experiment would be confounded. One of our goals is to control or eliminate such distractions.

Maturation This label seems to refer to experimental participants growing older during the course of an experiment, but that meaning is somewhat misleading. Although true maturation is possible during a longitudinal study, Campbell (1957) used the term **maturation** to refer to systematic time-related changes, but mostly of a shorter duration than we would typically expect from the term. Campbell and Stanley (1966) gave examples of participants growing tired, hungry, or bored through the course of an experiment. As you can see, maturation would be possible in an experiment that extends over some amount of time. How much time must occur before maturational changes can take place could vary widely depending on the experiment's demands on the participants, the topic of the experiment, or many participant variables (e.g., the participants' motivation, how much sleep they got the night before, whether they have eaten recently, and so on). Maturational changes are more likely to occur in an experiment that uses repeated measurements of the DV because experimental participants are involved in the research for a longer period of time. You should attempt to safeguard against such changes if your experiment requires long sessions for your participants.

Maturation A threat to internal validity; refers to changes in participants that occur over time during an experiment; could include actual physical maturation or tiredness, boredom, hunger, and so on.

Testing A threat to internal validity that occurs because measuring the DV causes a change in the DV.

Practice effect A beneficial effect on a DV measurement caused by previous experience with the DV.

Testing **Testing** is a definite threat to internal validity if you test participants more than once. As Campbell (1957) noted, if you take the same test more than once, scores on the second test may vary systematically from the first scores simply because you took the test a second time. This type of effect is known as a **practice effect**—your score does not change because of anything you have done other than repeating the test.

For example, the people who run the testing services for the two major college entrance exams (ACT and SAT) acknowledge the existence of testing effects when they recommend that you take their tests two or three times in order to achieve your maximum possible score. It appears that there is some “learning to take a college entrance exam” behavior that contributes to your score on such tests. This factor is separate from what the test is designed to measure (your potential for doing college work). Thus, those in charge are not surprised when your scores increase a little when you take the exam a second time; they actually expect an increase simply because you are familiar with the test and how it is given. However, two points about this specific testing effect are important to note. First, the effect is short-lived.

You can expect to gain some points on a second (and perhaps third) attempt, but you will not continue to gain points indefinitely just from the practice of taking the test. If improvement were constant, students could “max out” their scores merely by taking the test many times. Second, the expected improvement is small. Both the ACT and SAT administrators have a range of expected improvement that they consider normal. If you retake one of their tests and improve more than this normal range predicts, it is possible that they will question your score. You may be asked to provide documentation that accounts for your surprising improvement. If this documentation is not convincing, you may be asked to retake the test under their supervision. Unfortunately, the pressure to obtain scholarships (academic as well as athletic) has led some people to try to improve their scores through cheating or other unethical means.

In his discussion of testing effects, Campbell (1957) warned us against using **reactive measures** in our research. A measurement that is reactive changes the behavior in question simply by measuring it (see Chapter 4). Let us refresh your memory with an example. One popular topic for social psychology research is attitude change. If you are going to conduct an experiment on attitudes and attitude change, how will you measure your participants’ attitudes? If you are like most people, you probably responded, “With an attitude questionnaire.”

Reactive measures DV measurements that actually change the DV being measured.



Although questionnaires are commonly used to measure attitudes, they have a significant drawback. Can you figure out what that drawback is? Hint: This problem is particularly evident when you want to use an IV designed to change people’s attitudes. Can you think of a means for overcoming this difficulty?

Many attitude questionnaires are reactive measures. If we ask you a number of questions about how you feel about people of different racial groups, or about women’s rights, or about the president’s job performance, you can probably figure out that your attitude is being measured. The problem with knowing that your attitude is being measured is that you can alter your answers on a questionnaire in any way you wish. For example, if a woman administers a questionnaire to you, you may give answers that make you appear more sympathetic to women’s issues. If the questioner is of a different race than you, you may guard against giving answers that would make you look prejudiced. This type of problem is particularly acute if we use an experimental manipulation designed to change your attitude. If we give you a questionnaire about women’s rights, then show you a pro-women’s rights movie, and finally give you the questionnaire again, it doesn’t take a rocket scientist to figure out that we’re interested in finding out how the movie changed your attitude toward women’s rights! What can you do to avoid this problem in your research?

Campbell (1957) advocated using **nonreactive measures** that do not alter the participant’s response by virtue of measuring it. Psychologists (and others) have devised many tools and techniques for nonreactive measures: one-way mirrors, hidden cameras and microphones, naturalistic observation, deception, and so on (see Chapter 4). If we are interested in your attitude about women’s rights, perhaps we should see whether you attend

Nonreactive measures DV measurements that do not influence the DV being measured.

women's rights lectures or donate money to women's issues, ask whether you voted for a political candidate who supports women's rights, or measure some other such behavior. In other words, if we can obtain some type of behavioral measure that is harder to fake than a questionnaire response, we may get a truer measure of your attitude. As a psychology student doing research, you should guard against using reactive measures, which are often appealing because they seem so easy to use.

Instrumentation (Instrument Decay) As with maturation, Campbell's category of instrumentation has a broader meaning than is implied by its name. Part of Campbell's (1957) description of **instrumentation** referred to changes in measurement made by various apparatuses, although his list now seems quite outdated (e.g., "the fatiguing of a spring scales, or the condensation of water vapor in a cloud chamber," p. 299). He also included human observers, judges, raters, and coders in this category. Thus, the broad definition of instrumentation refers to changes in measurement of the DV that are due to the measuring "device," whether that device is an actual piece of equipment or a human.

This measurement problem related to equipment is probably less of an issue today than it was in "the good old days." (By the way, "the good old days" are probably defined as whenever your professor was in school.) We're sure your instructor has a horror story about the time the equipment broke down in the middle of an important day of data collection. The reality of the situation is that even today equipment can break down or develop flaws in measurement. You should check out your equipment before using it to collect data each day. One of the drawbacks of modern equipment, such as computers, is that it may break down, but that fact will not be obvious to you. Although you can certainly tell whether a computer is working, you may not be able to tell whether it appears to be working but actually has a bug or a virus. A careful experimenter checks out the equipment each day.

The same principle should govern your use of humans as monitors or scorers in your experiment. Often, you may serve as both the experimenter and the monitor or scorer for your experiment. This dual role places an extra obligation on you to be fair, impartial, and consistent in your behavior. For example, if you are scoring responses from your participants, you should have a set answer key with any judgment calls you might face worked out beforehand. So that you will not unconsciously favor one group over the other, you should score the responses blind as to a participant's group. If you have other persons scoring or monitoring your experiment, you should check their performance regularly to ensure that they remain consistent. Researchers often calculate interrater reliability (see Chapter 4) to determine whether instrumentation is a problem in a research project.

Statistical Regression You may remember the concept of statistical regression from your statistics course. When you deal with extreme scores, it is likely that you will see *regression to the mean*. This concept merely means that if you remeasure participants who have extreme scores (either high or low), their subsequent scores are likely to regress or move toward the mean. The subsequent scores are still likely to be extreme in the same direction, but not as extreme. For example, if you made a very high A (or very low F) on the first exam, your second exam score will probably not be as high (or as low). If a 7'0" basketball player has a son, the son will probably be tall but will probably be shorter than 7'0". If a 4'6" jockey has a son, the son will probably be short, but taller than 4'6". When you have extreme scores, it is simply difficult to maintain that high degree of extremity over repeated measures.

Instrumentation A threat to internal validity that occurs if the equipment or human measuring the DV changes the measuring criterion over time.

How does **statistical regression** play a role in internal validity? It is fairly common to measure participants before an experiment in order to divide them into IV groups (based on whatever was measured). For example, imagine an experiment designed to test the effects of a study course for college entrance exams. As a preliminary measure, we give an ACT or SAT test to a large group of high school students. From this pool of students, we plan to select our experimental participants to take the study course. Which students are we likely to select? The ones, of course, who performed poorly on the test. After they take the study course, we have them retake the entrance exam. More than likely, we will find that their new scores are higher than the original scores. Why is this? Did the study course help the students improve their scores? It is certainly possible; however, isn't it also possible that regression to the mean is at work? When you score near the low end of a test's possible score, there's not room to do much except improve. Sometimes it is difficult for people to see this point—they are convinced that the study course *must* be responsible for the improvement. To help illustrate the power of statistical regression, imagine a similar experiment performed with students who made the highest scores on the entrance exams. We want them to take our study course also. After our study course, we find that their second test scores are actually lower than the first set! Did our study course lower their scores? We certainly hope not. Having made extremely high scores on the first test, students find it difficult to improve on the second.

The advice concerning statistical regression should be clear. If you select participants on the basis of extreme scores, you should beware of regression to the mean as a possible explanation for higher or lower scores on a repeated (or similar) test.

Selection It should come as no surprise to you that **selection** can serve as a threat to internal validity. Before we conduct our experiment, it is imperative that we can assume that our selected groups are equivalent. Starting with equal groups, we treat them identically except for the IV. Then if the groups are no longer equal after the experiment, we can assume that the IV caused that difference.

Suppose, however, that when we select our participants, we do so in such a way that the groups are *not* equal before the experiment. In such a case, differences that we observe between our groups after the experiment may actually reflect differences that existed *before* the experiment began. On the other hand, differences after the experiment may reflect differences that existed *before* the experiment plus a treatment effect. Group differences that existed before the experiment are an extraneous variable, and the experiment is confounded.

Be careful not to confuse the selection problem with assignment of participants to groups. Selection typically refers to using participants who are already assigned to a particular group by virtue of their group membership. Campbell and Stanley (1966) cited the example of comparing people who watched a particular TV program to those who did not. People who choose to watch a program or not to watch a program probably differ in ways other than the knowledge of the program content. For example, it is likely that there are differences in people who watch and those who do not watch soap operas that are unrelated to the actual content of the soap operas. Thus, using soap opera viewers and non-soap opera viewers as groups to study the impact of a particular soap opera episode dealing with rape is probably not the best way to study the impact of that episode.

Statistical regression A threat to internal validity that occurs when low scorers improve or high scorers fall on a second administration of a test solely as a result of statistical reasons.

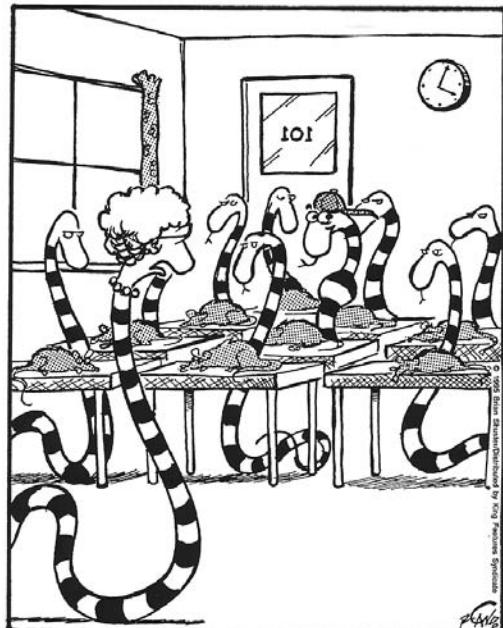
Selection A threat to internal validity that can occur if participants are chosen in such a way that the groups are not equal before the experiment; the researcher cannot then be certain that the IV caused any difference observed after the experiment.

Mortality As you might guess, “death” is the original meaning of *mortality* in this context. In research that exposes animals to stress, chemicals, or other toxic agents, mortality could actually occur in significant numbers. Mortality could become a threat to internal validity if a treatment were so severe that significant numbers of animals in the treatment group died. Simply because they had survived, the remaining animals in the treatment group would probably be different in some way from the animals in the other groups. Although the other groups would still represent random samples, those in the particular treatment group would not.

Mortality A threat to internal validity that can occur if experimental participants from different groups drop out of the experiment at different rates.

In human research, **mortality** typically refers to experimental dropouts. Remember the ethical principle from Chapter 2 that gives participants the right to withdraw from an experiment at any point without penalty? A situation in human research that is somewhat analogous to the animal example in the preceding paragraph could occur if many participants from one group withdraw from the experiment compared to small numbers from other groups. If such a differential dropout rate occurs, it is doubtful that the groups of participants would still be equal, as they were before the experiment began.

If your experiment contains a treatment condition that is unpleasant, demanding, or noxious in any way, you should pay careful attention to the dropout rate for the various groups. Differential dropout rates could also be a significant problem in research that spans a long time (weeks or months, for example). If one group shows a higher dropout rate, your



We certainly hope that you never experience this type of mortality in your research!

Reprinted with special permission of King Features Syndicate

“Today is the final session of our lab on rats, and Willy, I don’t think I have to remind you that this time if you eat yours, I’m not giving you another one.”

experiment may lack internal validity. One final word of caution: It is possible that groups in your experiment will not show different rates of dropping out but will have previously experienced different rates that helped compose the groups. Campbell (1957) cited the example of a very common experiment in educational settings, namely, making comparisons of different classes. Suppose we wish to determine whether our college's values-based education actually affects the values of the students. We choose a sample of freshmen and a sample of seniors and give them a questionnaire to measure their values in order to see whether their values are in line with our general education curriculum's values goals.



How might previous dropout rates damage the internal validity of this experiment?

In this experiment it should not surprise you to find that the seniors' values are more in line with our college's stated values. Although this effect *could* be a result of taking our values-based curriculum, it could also represent the effect of mortality through dropouts. Most colleges bring in large numbers of freshmen and graduate only some percentage of those students a few years later. What types of students are more likely to leave? Students whose values are not in line with the college's values are more likely to drop out than those who agree with the college's values. Thus, when we assess seniors' values, we are measuring those students who did *not* drop out.

Interactions with Selection **Interactions with selection** can occur when the groups we have selected show differences on another variable (i.e., maturation, history, or instrumentation) that vary systematically by groups. Let us give you an example for clarification. Suppose we are conducting a language development study in which we select our two groups from lower- and middle-class families. We wish to avoid selection as a threat to internal validity, so we pretest our children with a language test to ensure that our groups are equivalent beforehand. However, this pretesting could not account for the possibility that these two groups may show different maturation patterns, despite the fact that they are equal before the experiment begins. For example, if we look at children around the age of 1, we may find that the two classes of children show no language differences. However, by age 2, middle-class children might have larger vocabularies, talk more often, or have other linguistic advantages over lower-class children. Then if we look at children who are age 6, these linguistic advantages may disappear. If this were the case, comparing lower- and middle-class children's language proficiency at age 2 would show a selection–maturation interaction that would pose a threat to internal validity. Although one group of children would show an advantage over another group, the difference would not be a reliable, lasting difference. Therefore, any conclusion we reached that implied the permanence of such an effect would be invalid because the difference would have been based on different maturation schedules of the two groups.

Interactions with selection Threats to internal validity that can occur if there are systematic differences between or among selected treatment groups based on maturation, history, or instrumentation.

Similar interaction effects could occur between selection and history or between selection and instrumentation. A selection–history interaction that would jeopardize internal validity could occur if you chose your different groups from different settings—for example, different countries, states, cities, or even schools within a city. This problem is especially acute in cross-cultural research (see Chapter 7). These different settings allow for different local histories (Cook & Campbell, 1979) that might affect the outcome variable(s). The selection and history are confounded together because the groups were selected on a basis that allows different histories. An example of a selection–instrumentation interaction (based on human “instrumentation”) would be using different interpreters or scorers for participants who come from two different countries, assuming the IV was related to the country of origin.

Diffusion or imitation of treatment A threat to internal validity that can occur if participants in one treatment group become familiar with the treatment of another group and copy that treatment.

Diffusion or Imitation of Treatment **Diffusion or imitation of treatment** creates a problem with internal validity because it negates or minimizes the difference between the groups in your experiment. This problem can occur easily if your treatment involves providing some information to participants in one group but not to participants in another group. If the informed participants communicate the vital information to the supposedly uninformed participants, then the two groups may behave in identical manners.



If these two groups behave identically, what experimental conclusion would you draw? What would be wrong with this conclusion?

If the participants behave identically, the experimental outcome would indicate no difference between the groups based on the IV. The problem with this conclusion is that, in truth, there would actually be no IV in the experiment because the IV was supposed to have been the difference in information between groups. To avoid this threat, it is typical for experimenters to request that participants not discuss details of the experiment until after it is completed.

Experiments dealing with learning and memory are particularly susceptible to the imitation of treatments. For example, suppose we want to teach students in our classes a new, more effective study strategy. To simplify the experiment, we will use our 9:00 A.M. General Psychology class as a control group and our 10:00 A.M. General Psychology class as the experimental group. We teach the 9:00 A.M. class as usual, saying nothing to the students about how they should study for their quizzes and exams. In the 10:00 A.M. class, however, we teach our new study technique. After the first few quizzes and the first major exam, the students in the 10:00 A.M. class are so convinced of the effectiveness of their new study approach that they talk to their friends in the 9:00 A.M. class about it. Soon everyone in the 9:00 A.M. class knows about the new study technique and is also using it faithfully. At this point, we have the problem of treatment diffusion. Our participants in both groups are now using the same study strategy, which effectively no longer gives us an IV. It should be obvious that if you cannot guarantee control over the IV, the hope of having an experiment with internal validity is not strong.

Protecting Internal Validity

There are two approaches you could take to fight the various threats to internal validity. First, you can (and should) implement the various control procedures that we discussed in Chapters 6 and 7. These controls were specifically developed to deal with such problems. The second approach involves the use of a standard procedure. Just as detectives use standard police procedures to help them protect their cases, experimenters use standard research procedures, called experimental designs (see Chapters 10–12), to help ensure internal validity. We will have more to say about several experimental designs in subsequent chapters.

How important is internal validity? *It is the most important property of any experiment.* If you do not concern yourself with the internal validity of your experiment, you are wasting your time. Experiments are intended to derive cause-and-effect statements—to conclude that *X* causes *Y* to occur. You must take care to control any extraneous variables that might affect your dependent variable. You cannot count on your statistical tests to provide the necessary control functions for you. Statistical tests merely analyze the numbers you bring to the test—they do not have the ability to remove confounding effects (or even to discern that confounding has occurred) in your data.

■ REVIEW SUMMARY

1. For your experiment to have **internal validity**, it must be free from confounding owing to extraneous variables.
2. There are nine specific threats to internal validity against which we must guard.
3. The **history** threat can occur if meaningful events that would affect the DV occur between two measurements of the DV (repeated measures).
4. The **maturity** threat may occur if experimental participants show some change owing to time passage during an experiment.
5. The **testing** threat results if participants respond differently on the second measure of a DV simply because they have previously responded to the DV.
6. An **instrumentation** threat can be caused by equipment changes or malfunctions or to human observers whose criteria change over time.
7. **Statistical regression** is a threat that can occur when participants are selected because of their extreme scores on some variable. Extreme scores are likely to regress toward the mean on a second measurement.
8. The **selection** threat results if we select groups of participants in such a way that our groups are not equal before the experiment.
9. **Mortality** becomes a threat if participants drop out of the experiment at different rates for different conditions.
10. **Interactions with selection** can occur for history, maturation, or instrumentation. If we select participants in such a way that they vary on these dimensions, then internal validity is threatened.
11. **Diffusion** or **imitation of treatment** is a threat if participants learn about treatments of other groups and copy them.

Check Your Progress

1. Changing a behavior simply by measuring it refers to
 - a. a reactive measure
 - b. a measured IV
 - c. a manipulated IV
 - d. history
2. Comparing intact groups is likely to cause a problem because of
 - a. mortality
 - b. selection
 - c. maturation
 - d. history
3. Why is it important to evaluate your experiment for internal validity?
4. Match the internal validity threat with the appropriate situation.
 1. history
 2. maturation
 3. testing
 4. instrumentation
 - A. Your DV scorer gets sick and you recruit a new person to help you.
 - B. You are conducting an experiment on racial prejudice and a race riot occurs between tests.
 - C. Your participants grow bored and disinterested during your experimental sessions.
 - D. You use a before-and-after DV measurement and the participants remember some of their answers.
5. Match the threat to internal validity with the correct situation.
 1. statistical regression
 2. mortality
 3. selection
 4. diffusion of treatments
 - A. Many participants find one treatment condition very boring and quit.
 - B. You choose boys from lower-class homes and girls from upper-class environments.
 - C. Students in your control group talk to students in the experimental group and imitate their treatment.
 - D. You select the worst students in the class and try a new tutoring strategy.
6. You want to compare the formal education of college students and senior citizens. You select a group of each type of participant and give each a written test of math, social studies, and grammatical information. What threat of internal validity appears likely in this situation? Why?
7. Experimental dropouts are a problem linked to
 - a. mortality
 - b. selection
 - c. maturation
 - d. history

External Validity: Generalizing Your Experiment to the Outside

In the first part of this chapter we covered the concept of internal validity, which is one way you must evaluate your experiment. As you will remember, internal validity concerns the question of whether your experiment is confounded: Is your IV the only possible explanation for your finding? In this section we will cover another type of validity you should assess in your experiment.

The second type of evaluation that you must make of your experiment involves **external validity**. When you consider external validity, you are asking a question about **generalization**. The root word of *generalization* is *general*. Of course, *general* is essentially the opposite of *specific*. So, when we want to generalize our experimental results, we wish to move away from the specific to the more general. In essence, we would like to take our results beyond the narrow confines of our specific experiment. Imagine that you have conducted an experiment using 20 college students as your participants. Did your experimental question apply only to those 20 college students, or were you attempting to ask a general question that might be applicable to a larger group of students at your school, or to college students in general, or perhaps even to people in general? Few experimenters gather data that they intend to apply only to the participants they actually use in their experiments. Instead, they hope for general findings that will apply more broadly. When you examine different definitions of psychology, you find that many include a phrase about psychology being the “science of behavior.” We don’t think you’ve ever seen a definition that referred to the “science of blondes’ behavior,” or “the science of Californians’ behavior,” or “the science of children’s behavior”—probably not even “the science of human behavior.” Psychologists attempt to discover information that applies to everyone and everything. As Sherlock Holmes said, “One’s ideas must be as broad as Nature if they are to interpret Nature” (Doyle, 1927, p. 37). When we discover principles that apply broadly, then we have found behavior that is lawful. Finding such principles is a tall order, especially for a single experiment, but we keep this goal in mind as we do our research. We are ever mindful of the need and desire to extend our findings beyond our specific experiment.

Let us share with you a dilemma that this chapter posed for us. The dilemma revolved around where to put this chapter in the text. By placing it where we have, before we consider experimental designs, you may get the idea that you have to worry about evaluating your experiment only for external validity before you design and complete your experiment. Nothing could be further from the truth. Rather, we hope that you are constantly evaluating your experiment. Much of what we have written for this chapter is designed to help you evaluate your experiment from the day you begin planning your methodology all the way through the completion of your project. When designing their experiments, researchers take precautions that will allow them to increase the generalizability of their final results. Our dilemma, then, was whether to place this chapter nearer the beginning of the book or nearer the back. We chose to put it nearer the front to alert you to this issue that merits consideration throughout the entire course of conducting a research project.

External validity A type of evaluation of an experiment; do the experimental results apply to populations and situations that are different from those of the experiment?

Generalization Applying the results from an experiment to a different situation or population.

After you have satisfied yourself that your experiment (or an experiment that you are evaluating) is internally valid, then you can focus on its potential for external validity. As we mentioned, the question of external validity revolves around whether you can take the results of a particular experiment and generalize them beyond that original experiment. Generalization is an important aspect for any science—there are very few occasions when scientists wish to restrict their conclusions solely to the relevant conditions of their original experiment. Although police detectives are usually interested in a specific culprit, they are also interested in developing criminal profiles that describe general traits of a specific type of offender (e.g., the profile of mass murderers or of child molesters).

Population generalization Applying the results from an experiment to a group of participants that is different and more encompassing than those used in the original experiment.

There are three customary types of generalization in which we are interested. In **population generalization**, we are concerned about applying our findings to organisms beyond our actual experimental participants. When our students conduct research with fellow students as participants, they are not interested in applying their results only to the students at their respective colleges. We doubt that many of you who do experimental projects will try to answer a question only about students at your school. Likewise, psychologists are not typically concerned with trying to discover truths only about the humans or the animals who participate in their studies. In all the cases we have just mentioned, the researchers (including you) are trying to discover general principles about behavior that are applicable to people or animals as a whole. Thus, we must concern ourselves with whether our results will apply to a larger group than our experimental participants.

Environmental generalization Applying the results from an experiment to a situation or environment that differs from that of the original experiment.

Environmental generalization refers to the question of whether our experimental results will apply to situations or environments that are different from those of our experiment. Again, it is doubtful that we are interested in finding results that are specific to the environment in which we tested our participants. Can we generalize our results from the college classroom or research lab in which we collected our data to other classrooms or labs or even the real-world environments in which behavior naturally occurs? For example, if students learn a list of words in a laboratory, does that behavior have any relevance to the real-world behavior of studying for an exam?

Temporal generalization Applying the results from an experiment to a time that is different from the time when the original experiment was conducted.

Finally, **temporal generalization** describes our desire that our research findings apply at all times and not only to a certain time period. The discovery of seasonal affective disorder (depression which occurs only during winter months) is, of course, an exception to generalized information about depressive disorders. Another example of research findings that might be specific to a certain time are the many findings of gender differences from research in the 1960s. Today, after several decades of progress on women's rights, we should wonder whether the

1960s findings are still applicable today. Such an issue raises thorny questions for the concept of temporal generalization.

We must ask questions about the external validity of our research results because of the very property that we mentioned in the first paragraph of this chapter: *internal validity*. To achieve a high degree of internal validity, researchers seek to exert control over a large number of factors. The best way to exert control over factors is to conduct your experiment in a lab (or similar setting) with participants who are highly similar.



Reread the last sentence. Explain why this statement is true. Can you figure out why exerting control, which helps us in terms of internal validity, ends up weakening our external validity?

When we conduct our research in a laboratory, we can control many factors that might function as extraneous variables in the real world. For example, a researcher can control such factors as temperature, lighting, and outside noise. In addition, it is possible to control the time when the participants come to the experimental situation, as well as when and under what conditions they experience the IV. In short, the lab allows you to remove or hold constant the effects of many extraneous variables, which would be virtually impossible to do in the real world. In a similar fashion, using participants who are highly alike helps reduce nuisance variation (see our discussion of nuisance variables in Chapter 6). For example, if you use only college students as participants, you will have an idea of their general reading ability if you have to compose a set of instructions for them to read. In the real world, potential participants would have a much wider range of reading abilities, thus adding nuisance variation to the experiment.

We hope that you can see that these advantages for internal validity become disadvantages for external validity. When we control so many factors so tightly, it is difficult to know whether our results will apply to people or animals who are markedly different from our research participants and who encounter our research factors in the real world rather than in the artificial laboratory.

Threats to External Validity (Based on Methods)

Donald Campbell, who wrote extensively about internal validity, has also written widely about factors relating to external validity. We will summarize four factors affecting external validity from Campbell and Stanley (1966).

Interaction of Testing and Treatment When Campbell and Stanley (1966) rated three experimental designs for internal validity (pretest–posttest control group, Solomon four-group design, and posttest-only control group—see Chapter 13), they also attempted to rate them concerning external validity (although that task is much more difficult). Because we will make several references to Chapter 13 in the following discussion, you may wish to skim over it at this point. The **interaction of testing and treatment** is the most obvious threat to external validity, and it occurs for the pretest–posttest control group design (see Figure 8-1), which we will discuss in Chapter 13. External validity is threatened because both groups of participants are pretested and there is no control to determine whether the pretesting has had an effect. As Solomon (1949) pointed out, “There is a great possibility that merely taking a pre-test changes the subjects’ attitudes toward the training procedures” (p. 141).

Imagine that you are working on an experiment dealing with prejudice toward people with mental disorders, specifically people with schizophrenia. You decide to test your participants by giving them a scale that measures their prejudice toward people with schizophrenia

Interaction of testing and treatment A threat to external validity that occurs when a pretest sensitizes participants to the treatment yet to come.

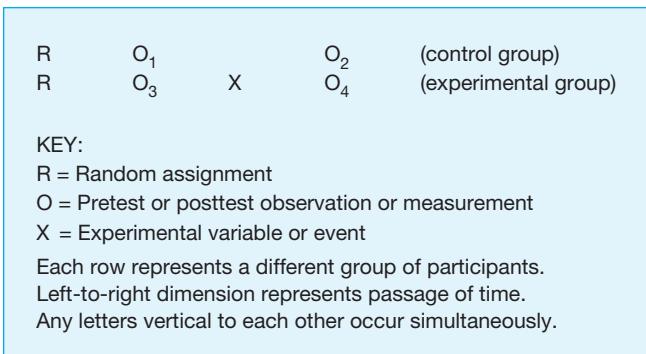


FIGURE 8-1 The Pretest–Posttest Control-Group Design.

before showing them a movie based on the book, *A Beautiful Mind* (Grazer & Howard, 2001), which depicted John Nash's struggles with schizophrenia before he won a Nobel Prize for his work in mathematics. Does it seem likely that these people's responses to the movie will be different from those of a group of people who were not pretested? This question is the essence of an interaction between testing and treatment: Because of a pretest, your participants' reaction to the treatment will be different. The pretest has a sensitizing effect on your participants; it is somewhat like giving a giant hint about the experiment's purpose before beginning the experiment. The pretesting effect is particularly troublesome for experiments that deal with attitudes and attitude change. This testing and treatment interaction is the reason why researchers developed nonpretesting designs such as the posttest-only control-group design (see Chapter 13). Also, although the Solomon four-group design (see Chapter 13) does incorporate pretests for two groups, it includes no pretesting of the remaining two groups, thus allowing for a measurement of any pretesting effect.

Interaction of Selection and Treatment You will remember we covered “interactions with selection” as a threat to internal validity (we will revisit it in Chapter 13). In those cases, selection potentially interacted with other threats to internal validity such as history, maturation, and instrumentation. In this instance, however, the threat to external validity consists of an interaction between selection and treatment.

An **interaction of selection and treatment** occurs when the effects that you demonstrate hold true only for the particular groups that you selected for your experiment. Campbell and Stanley (1966) noted that the threat of a selection–treatment interaction becomes greater as it becomes more difficult to find participants for your experiment. As it becomes harder to get participants, it becomes more likely that the participants you do locate will be unique and not representative of the general population.

One of the authors of this book encountered a selection–treatment interaction when he was attempting to complete his doctoral dissertation, in which he was testing a particular explanation of a common and easy-to-find phenomenon (better memory for material studied repeatedly at longer intervals rather than shorter intervals). Because of time pressures, he had to conduct his experiment during summer school. When he analyzed the data, the supposedly easy-to-find phenomenon was not evidenced by the statistics. Rather than

Interaction of selection and treatment A threat to external validity that can occur when a treatment effect is found only for a specific sample of participants.

concluding that the phenomenon did not exist (contrary to *much* published evidence), he decided that this finding was a result of a selection–treatment interaction. When he conducted his experiment a second time during the regular semester, the ubiquitous effect was found. It appears that, for some reason, the summer school students were a unique population.

Reactive Arrangements Reactive arrangements revolve around the artificial atmosphere of many psychological experiments. Because of our desire for control, we may create a highly contrived situation in a laboratory in which we attempt to measure a real-world behavior. There are some cases in which this type of arrangement is probably a dismal failure. As you saw in Chapter 7, experimental participants, particularly humans, seek to uncover clues about their environments. They may react to our experimental manipulations, but they also may react to subtle cues they find in the experimental situation. As Campbell and Stanley (1966) remarked, “The play-acting, outguessing, up-for-inspection, I’m-a-guinea-pig, or whatever attitudes so generated are unrepresentative of the school setting, and seem to be qualifiers of the effect of X, seriously hampering generalization” (p. 20). Although their comment was directed at educational research, if we substitute “real world” for “school setting,” the quotation remains valid. Thus, **reactive arrangements** refer to those conditions of an experimental setting (other than the IV) that alter our participants’ behavior. We cannot be sure that the behaviors we observe in the experiment will generalize outside that setting because these artificial conditions do not exist in the real world.

In Chapter 4 we mentioned a series of studies that are usually cited as classic examples of reactive arrangements: the Hawthorne studies. You will recall that researchers studied productivity of a subset of plant workers over a period of time. This subset of workers tended to increase productivity regardless of the changes instituted in their environment. One of the representative experiments dealt with the level of illumination in a work area. A control group worked in a room with 10 foot-candles of light, whereas the experimental group’s illumination began at 10 foot-candles and was gradually decreased by 1 foot-candle at a time until it reached 3 foot-candles. At this point, “the operatives protested, saying that they were hardly able to see what they were doing, and the production rate decreased. The operatives could and did maintain their efficiency to this point in spite of the discomfort and handicap of insufficient illumination” (Roethlisberger & Dickson, 1939, p. 17).

In Chapter 7 we introduced the concept of **demand characteristics**. According to Orne (1962), demand characteristics can convey the experimental hypothesis to the participants and give them clues about how to behave. You (or your friends) may have experienced something similar to demand characteristics if you are a psychology major. When you tell people you are a psychology major, a common response is “Are you going to psychoanalyze me?” Rather than responding to you as an individual, people are responding to the demand characteristics of what they believe about psychology majors.

According to Orne, it is impossible to design an experiment without demand characteristics. He also believed that demand characteristics make generalizations difficult because it is not clear from a set of research findings whether the participants are responding to an experiment’s IV, its demand characteristics, or both. It seems that reactive arrangements tend to increase demand characteristics, thus creating further difficulties with generalization.

Reactive arrangements
A threat to external validity caused by an experimental situation that alters participants’ behavior, regardless of the IV involved.

Demand characteristics
Features of the experiment that inadvertently lead participants to respond in a particular manner.

Multiple-treatment interference

A threat to external validity that occurs when a set of findings results only when participants experience multiple treatments in the same experiment.

Multiple-Treatment Interference As you can probably guess from the name of this threat to external validity, **multiple-treatment interference** can occur only in experiments that involve presenting more than one treatment to the same participants (repeated-measures designs). The potential problem that arises in repeated-measures designs is that the findings we obtain may be specific to situations in which the experimental participants experienced these multiple treatments. If they received only one treatment, the experimental results might be different.

Campbell and Stanley (1966) cited Ebbinghaus's memory work as an example of multiple-treatment interference. Ebbinghaus, a pioneer in the field of verbal learning, conducted many experiments in learning lists of nonsense syllables by serving as his own experimental participant. He is responsible for giving us early versions of many common learning phenomena such as learning and forgetting curves. As it turns out, however, some of Ebbinghaus's findings are specific to people who have learned a large number of lists of nonsense syllables and do not apply to people learning a single list of nonsense syllables. Thus, Ebbinghaus's results show a clear failure to generalize, based on the concept of multiple-treatment interference.

This ends the list of four threats to external validity summarized by Campbell and Stanley (1966). Let's take a brief look at five additional threats to external validity that are based on our experimental participants.

Threats to External Validity (Based on Our Participants)

We must always remember that our experimental participants are unique individuals, just as the detective must remember that fact about each criminal. Certain unique traits may get in the way of our ability (or the detective's ability) to draw general conclusions.

Comparative psychology

The study of behavior in different species, including humans.

The Infamous White Rat In 1950 Frank Beach sounded the alarm about a distressing tendency he noticed in **comparative psychology**. By charting the publication trends in that field from 1911 to 1947, he found that more and more articles were being published but fewer species were being studied. The Norway rat became a popular research animal in the 1920s and began to dominate research in the 1930s. Beach examined 613 research articles and found that 50% of them dealt with the rat, despite the fact that the Norway rat represented only 0.001% of all living creatures that could be studied (Beach, 1950). In a follow-up study, Smith, Davis, and Burleson (1995) found that articles published between 1993 and 1995 showed attention to diversified species, but still rats and pigeons accounted for more than 75% of the studies published in the *Journal of Experimental Psychology: Animal Behavior Processes and Animal Learning & Behavior* during those years.



Why should Beach's data (and the follow-up study) give us reason for concern regarding external validity?

Two obvious concerns with external validity arise from these data. First, if you are interested in the behavior of subhumans, generalizing from rats (and pigeons) to all other animals may be a stretch. Second, if you are interested in generalizing from animal to human behavior, there are certainly closer approximations to humans than rats (and pigeons) in the animal kingdom.



First published in the APA Monitor. Reprinted by permission of Warren R. Street.

The Ubiquitous College Student We hope that you have conducted a literature search at this point in your course. If you have, you should have noticed that experimental articles contain a "subjects" or "participants" subsection within the method section (see Chapter 14). (The label "participants" replaced "subjects" in the fourth edition of the American Psychological Association's *Publication Manual*, 1994.)

If your literature search involved human participants, you probably noticed that college students very likely served as the participants in the research you reviewed. There is a simple explanation for this fact. Psychology departments that support animal research typically have animal lab facilities—often colonies of rats, as we saw in the preceding section. Psychologists who want to conduct human research have no such labs filled with participants, of course, so they turn to a ready, convenient source of human participants—students enrolled in introductory psychology courses (a technique referred to as **convenience sampling**). Often students may be required to fulfill a certain number of hours of experimental participation in order to pass their introductory class. As well as providing a large supply of participants for psychology experiments, this requirement should be an educational experience for the students. Such requirements have generated heated debates, some of which have focused on ethics. You will note in the ethical principles covered in Chapter 2 that coercion is not an acceptable method of finding experimental participants.

Convenience sampling
A researcher's sampling of participants based on ease of locating the participants; often it does not involve true random selection.

We are including this section not to focus on the ethics of using students as research participants but rather to focus on the implications of this reliance. Sears (1986) raised this very issue with regard to social psychology. He pointed out that early social psychologists studied a variety of types of participants, including populations as diverse as radio listeners, voters, soldiers, veterans, residents of housing projects, industrial workers, union members, and PTA members. However, by the 1960s psychology had become entrenched in using laboratory experiments conducted with college students as participants.

Is it a problem to rely almost exclusively on college students? Clearly, such reliance is similar to the reliance on the white rat by animal researchers. Sears (1986) cited evidence concerning the differences between adolescents and adults and between college students and other late adolescents. He worried that several notable social psychology findings might be the product of using college students as participants rather than the larger population in general.



Reread the last sentence. Can you identify Sears's position as being consistent with one of Campbell and Stanley's threats to external validity?

Sears's hypothesis that some social psychology findings may be specific to college students fits into Campbell and Stanley's selection-treatment interaction. If Sears is correct, then a specific treatment might "work" only for college students. In addition, a particular IV might have a stronger (or weaker) effect for college students than for the general population.

As an example, Sears mentioned that much of social psychology revolves around attitudes and attitude change: The research often shows that people's attitudes are easily changed. Developmental research, however, indicated that the attitudes of late adolescents and young adults are less crystallized than those of older adults. Thus, we have to wonder whether people's attitudes really are easily changed or whether this finding is merely an artifact of the population that is most often studied: the college student.

Although Sears concentrated on social psychology findings, it does not take much thought to see how similar problems might exist in other traditional research areas of psychology. For example, because college students are both students and above average in intelligence, might their learning and memory processes be a poor representation of those processes in the general population? Motivational patterns in adolescents are certainly different in some ways from those processes in adults. The stresses affecting college students differ from those facing adults—might their coping processes also be different? We know that some mental illnesses are age related and therefore strike older or younger people more often. Overall, we should be careful in generalizing to the general population from results derived solely from college students.

The "Opposite" or "Weaker" or "Inferior" or "Second" Sex All four of these derogatory labels—*opposite*, *weaker*, *inferior*, and *second*—have been applied to women at various points in time. As you took your school's required history, literature, and humanities classes, you probably noticed that famous women were in short supply. For many years (some would argue still now), women have simply not had the same opportunities as men in many situations.

This supposed inferiority of women carried over into psychological theories. For example, Freud's theories are biased toward men—remember castration anxiety and the Oedipus complex? The “parallel” concepts for women of penis envy and the Electra complex seem almost to be afterthoughts. Feminists have repeatedly attacked the concept of penis envy, supposedly part of equating Freud's theory for women and men. Erik Erikson labeled his theory of psychosocial crises the “Eight Stages of Man.” Not only did the name not acknowledge women, but neither did the stages. Many women have not identified with Erikson's stages to the extent that men have. As Carol Tavris (1992) eloquently pointed out about developmental theories, “Theorists writing on adult development assumed that adults were male. Healthy ‘adults’ follow a single line from childhood to old age, a steady path of school, career, marriage, advancement, a child or two, retirement, and wisdom—in that order” (p. 37).

Tavris's thesis is that “despite women's gains in many fields in the last 20 years, the fundamental belief in the normalcy of men, and the corresponding abnormality of women, has remained virtually untouched” (p. 17). Tavris's ideas make us face an interesting question: Are we developing a body of knowledge that pertains to all organisms, regardless of sex? One detail we will emphasize about APA style in Chapter 14 is its requirement of nonsexist language in writing. Words carry hidden meanings, and sexist writing may lead to generalizations that are unwarranted. As psychology students, you should definitely be attuned to this issue. Surveys of psychology undergraduates across the country find that about three-quarters of all psychology majors are women. You may even encounter difficulty finding male participants for your future experiments. Without adequate male representation, the generality of your findings could be questioned.

Again, a word of caution is in order. When we generalize our findings, we typically generalize them to as large a segment of the population as we can. However, we should not generalize research conducted on only one sex to both sexes. Beware of drawing sexist conclusions from your research.

Even the Rats and Students Were White This section's heading is an adaptation of the title of Robert Guthrie's thought-provoking book *Even the Rat Was White* (1976). In this book, Guthrie chronicled many of the “scientific” attempts made to measure and categorize African Americans as inferior to Caucasians, including anthropometric measurements (e.g., skull size and cranial capacity), mental testing, and the eugenics movement (improvement of heredity through genetic control).

Some psychology students may assume that there were no African American pioneers in psychology, an assumption that is far from the truth. Guthrie (1976) listed 32 African Americans who earned doctorates in psychology and educational psychology in American universities between 1920 and 1950. He provided brief background and career sketches for 20 of these individuals. Let us give you a glimpse into two of these persons. Guthrie devoted a chapter to Francis Cecil Sumner, whom he labeled the “Father of black American psychology” (p. 175). Sumner, born in Pine Bluff, Arkansas, was the first African American to earn a Ph.D. in psychology in the Western Hemisphere. He earned his degree from Clark University in June 1920. His chief instructor was G. Stanley Hall; he also took courses from E. G. Boring. Sumner's dissertation, “Psychoanalysis of Freud and Adler,” was published in *Pedagogical Seminary* (later renamed the *Journal of Genetic Psychology*). Sumner went on to teaching positions at Wilberforce University and West Virginia Collegiate Institute. In 1928 he accepted a position at Howard University, where he became chair of the newly established department of psychology.

in 1930. Under Sumner's leadership, Howard became one of the major producers of African American Ph.D.s in psychology. Sumner remained at Howard until his death in 1954.

Ruth Winifred Howard was the first African American woman to earn a doctorate in psychology, from the University of Minnesota in 1934. (Inez Beverly Prosser had earned a Ph.D. in *educational psychology* from the University of Cincinnati one year earlier.) Howard's dissertation, "A Study of the Development of Triplets," was one of the first studies to deal with a sizable group of triplets. She completed a clinical internship at the Illinois Institute for Juvenile Research and spent her career in private practice in clinical psychology.

Just as history has failed to record the accomplishments of many women throughout time, it has largely ignored the accomplishments of African Americans and other minority groups. Jones (1994) pointed out the inherent duality of being an African American—that one is separately considered an African and an American. Other writers have made similar points about and pointed out problems for members of other minority groups; for example, Marín (1994) for Hispanics, Lee and Hall (1994) for Asians, and Bennett (1994) for American Indians. According to Lonner and Malpass (1994b), projections indicate that sometime in this century the United States will be 24% Hispanic, 15% African American, and 12% Asian; Caucasians will make up less than half the population for the first time in U.S. history. When we conduct research and make generalizations, we should be cautious that we do not exclude minority groups from our considerations.

Even the Rats, Students, Women, and Minorities Were American Although experimental psychology's early roots are based in western Europe, this aspect of the discipline quickly became Americanized, largely as a result of the influence of John B. Watson's behaviorism. For many years, the study of human behavior was actually the study of *American* behavior. In the mid-1960s, however, this imbalanced situation slowly began to change as psychologists started taking culture and ethnicity more seriously (Lonner & Malpass, 1994a). We mentioned this new emphasis on ethnicity in the previous section and wish to mention the notion of culture in this section.

The field of cross-cultural psychology has evolved from those changes that began in the 1960s. Today you can find textbooks and courses devoted to this topic. We introduced you to cross-cultural research in Chapter 7. Because cross-cultural psychology "tests possible limitations of our traditional knowledge by studying people of different cultures" (Matsumoto, 2000, p. 10), it is closely intertwined with external validity. In fact, cross-cultural psychology tests the very limits of external validity. It asks how far a set of experimental findings can be extended to other cultures. If psychology is made up of findings that are applicable only to Americans, then we should not claim that we know everything there is to know about *human behavior!* Making an ethnocentric claim such as that is akin to American professional baseball claiming to play the *World Series*, which actually began as a series of games sponsored by the *World* newspaper. What about the professional teams in Japan and other countries? Similarly, we shouldn't study maternal behaviors in the United States and claim that we know everything there is to know about mothering. Again, a word to the wise: Let's be careful about making grandiose claims about external validity when we don't have all the data. With our shrinking world today, it is becoming easier and easier to conduct true cross-cultural research. The books written by Lonner and Malpass (1994b) and Matsumoto (2000) will give you some good ideas about behavioral differences that exist as a function of culture.

The Devil's Advocate: Is External Validity Always Necessary?

Douglas Mook (1983) published a thought-provoking article titled "In Defense of External Invalidity." In this article, Mook attacked the idea that all experiments should be designed to generalize to the real world. He maintained that such generalization is not always intended or meaningful.

Mook cited as an example Harlow's work with baby rhesus monkeys and their wire-mesh and terry-cloth surrogate mothers. As you probably remember, the baby monkeys received nourishment from the mesh mothers and warmth and contact from the cloth mothers. Later, when faced with a threatening situation, the baby monkeys ran to the cloth mothers for comfort. As virtually all introductory psychology students learn, theorists and researchers have used Harlow's work to support the importance of contact comfort for the development of attachment. Was Harlow's research strong in external validity? Hardly! Harlow's monkeys could scarcely be considered representative of the monkey population because they were born in the lab and orphaned. Was the experimental setting lifelike? Hardly! How many baby monkeys (or human infants) will ever be faced with a choice between a wire or cloth "mother"?

Did these shortcomings in external validity negate Harlow's findings? As Mook pointed out, that depends on the conclusion that Harlow wished to draw. What if Harlow's conclusion had been, "Wild monkeys in the jungle probably would choose terry-cloth over wire mothers, too, if offered the choice" (Mook, 1983, p. 381)? Clearly, this conclusion could not be supported from Harlow's research because of the external validity problems mentioned in the previous paragraph.

On the other hand, Harlow *could* conclude that his experiment supported a theory of contact comfort for mother–infant attachment over a nourishment-based (drive-reduction) theory. Is there a problem with this conclusion? No, because Harlow did not attempt a generalization—he merely drew a conclusion about a theory, based on a prediction from that theory. Mook argued that our concern with external validity is necessary only when we are trying to "predict real-life behavior in the real world" (p. 381). Of course, there are reasons to conduct research that do not involve trying to predict behavior in the real world. Mook pointed out four alternative goals of research that do not stress external validity. First, we may merely want to find out whether something *can* happen (not whether it usually happens). Second, we may be predicting from the real world to the lab—seeing a phenomenon in the real world, we think it will operate in a certain manner in the lab. Third, if we can demonstrate that a phenomenon occurs in a lab's unnatural setting, the validity of the phenomenon may actually be strengthened. Finally, we may study phenomena in the lab that don't even have a real-world analogy.

We do not mean to undermine the importance of external validity by presenting Mook's arguments; however, it is important to know that not all psychologists worship at the shrine of external validity. Being concerned with real-world applications is certainly important, but (unlike internal validity) it is not an absolute necessity for every experiment. By the same token, the detective does not attempt to generate a generalized criminal profile for each crime that is committed.



Just as with internal validity, we have looked at nine different threats to external validity. We hope you have been considering these threats in light of your own potential experiment. What can you do to avoid external validity problems?

This is probably the most difficult Psychological Detective question in this chapter; there is a sense in which this question is a trick. It *may* be impossible to conduct a single experiment that has perfect external validity. If you attempted to devise an experiment that answered every threat to external validity that we listed, you might be old and gray before you completed the experiment. Imagine trying to find a pool of human experimental participants that would satisfy sex, race, ethnicity, and cultural generalizability. You would have to find large numbers of participants from around the world—clearly an impossible situation.

Does the dismal conclusion of the previous paragraph mean that we should merely throw up our hands and quit? Are we justified in ignoring the problem of external validity? Of course not! The important question becomes one of what steps we can take to maximize our chances of achieving external validity. First, we recommend that you pay particular attention to the first four external validity threats that we discussed. To a large extent, you do have control over interactions of testing or selection and treatment, reactive arrangements, and multiple-treatment interference. Careful experimental planning can usually allow you to avoid problems with those factors.

If you can control the methodological threats to external validity, that leaves you with the participant-related threats. It is unrealistic to expect any one experiment to be able to include all the various participants we would have to generalize our results across the world's population. Campbell and Stanley (1966) wrote that "the problems of external validity are not logically solvable in any neat, conclusive way" (p. 17). What is the solution?

Replication An additional scientific study that is conducted in exactly the same manner as the original research project.

It seems logical that the solution must involve an approach we discussed in Chapter 1: **replication**. When we replicate an experimental finding, we are able to place more confidence in that result. As we begin to see the same result time after time, we become more comfortable with the idea that it is a predictable, regularly occurring result. However, it is shortsighted for us to test continually the same types of participants in every experiment we conduct, whether those participants are white rats or White American college students. We must go beyond replication to experiments that involve replication with extension.

Replication with extension An experiment that seeks to confirm (replicate) a previous finding but does so in a different setting or with different participants or under different conditions.

In a **replication with extension** experiment, we retest for a particular experimental finding, but we do so in a slightly (or even radically) different context. For example, if we think we know a great deal about how American college students learn lists of nonsense syllables, we should not try to generalize those results to all humans. Instead, we should broaden our participant population. Do the same rules apply when we test elderly people, children, and less-educated people? Do the same findings occur when we test Hispanic Americans, Asian Americans, and African Americans? Are the results similar in Japan, China, Peru, Australia, and so on? When we begin to collect data from experiments as indicated by the previous three sentences, then we truly start learning whether our findings are generally applicable.

There is another advantage to experiments that use replication with extension. Many times pure replication experiments are met with disapproval. Journals rarely, if ever, publish a straight replication study. Many undergraduates who complete required experimental projects are told that their projects must be original, that the project cannot be a replication study. A replication with extension that involves enough new work and produces enough new information might be acceptable either for a class requirement or for publication—or for both if you're extremely fortunate!

A 2001 research study conducted by Carolyn Ann Licht, a student at Marymount Manhattan College in New York, and Linda Zener Solomon, her faculty advisor, illustrated a replication with extension experiment. Licht (2000) had previously compared stress levels of employees in two New York City organizations and found that stress levels were higher for employees of a nonprofit organization than for employees of a for-profit organization. Licht and Solomon extended the earlier study to include more organizations and more geographic areas. Their findings mirrored Licht's earlier findings, leading Licht and Solomon to conclude that the "results were consistent with previous findings and increased their external validity" (p. 14).

So what is the conclusion regarding external validity? We believe it is unrealistic to expect every (or any) experiment to be applicable to the entire world of animals or people. That type of external validity is probably a myth that exists only in fairy tales. Nonetheless, we also believe that you should strive to make your experiment as externally valid as possible: When you have choices, opt for the choices that will increase the external validity of your findings.

■ REVIEW SUMMARY

1. If your experiment is internally valid, then you can worry about **external validity**, which deals with applying your findings to new groups (**population generalization**), new situations (**environmental generalization**), and new times (**temporal generalization**).
2. There are many threats to external validity. The **interaction of testing and treatment** threat can occur if you pretest your participants and the pretest changes their reaction to the posttest.
3. The **interaction of selection and treatment** threat occurs if your findings apply only to the groups you selected for your experiment.
4. **Reactive arrangements** can threaten external validity if the experimental situation changes our participants' behavior.
5. **Multiple-treatment interference** may threaten external validity if participants experience more than one treatment and this multiple participation causes a change in behavior.
6. Experimental psychology has been criticized for using only a narrow slice of possible types of participants in research. Animal research is particularly imbalanced by its overuse of white rats as participants.
7. Human research has tended to focus on white American college students as participants. Many theories seem to be aimed heavily or exclusively at men.
8. For reasons of external validity, there is currently a greater focus on research involving participants from the general population, women, ethnic minorities, and other cultures.
9. Mook has pointed out that there are occasional instances when the external validity question is not relevant to research.
10. It is nearly impossible to devise a single experiment that can answer all questions about external validity. Conducting experiments that involve **replication with extension** is a good way simultaneously to increase external validity and gather new information.

Check Your Progress

1. What is external validity? Why is it important to psychology?
2. Generally speaking, as internal validity increases, external validity
 - a. increases
 - b. decreases
 - c. remains the same
 - d. fluctuates unpredictably
3. Distinguish among population, environmental, and temporal generalization.
4. Why should we be concerned about trying to use different types of participants (such as minorities and people who are not college students) in psychology studies?
5. What is cross-cultural psychology? Why is it of particular relevance to this section of the chapter?
6. Match the external validity threat with the proper description.
 1. testing-treatment interaction
 2. selection-treatment interaction
 3. reactive arrangements
 4. multiple-treatment interference
 - A. An effect occurs only if participants experience all the experiment's treatment conditions.
 - B. Women, but not men, demonstrate an experimental effect.
 - C. Only participants who are pretested demonstrate an experimental effect.
 - D. Demand characteristics provide cues to the participants about how they should respond.
7. Why did Mook argue that external validity is not always necessary?
8. Why is the notion of external validity for a single experiment virtually impossible to achieve?

Key Terms

Internal validity, 147	Interactions with selection, 153	Interaction of selection and treatment, 160
History, 147	Diffusion or imitation of treatment, 154	Reactive arrangements, 161
Maturation, 148	External validity, 157	Demand characteristics, 161
Testing, 148	Generalization, 157	Multiple-treatment interference, 162
Practice effect, 148	Population generalization, 158	Comparative psychology, 162
Reactive measures, 149	Environmental generalization, 158	Convenience sampling, 163
Nonreactive measures, 149	Temporal generalization, 158	Replication, 168
Instrumentation, 150	Interaction of testing and treatment, 159	Replication with extension, 168
Statistical regression, 151		
Selection, 151		
Mortality, 152		

Looking Ahead

So far, our view of research in psychology has been rather general. At this point in the book we are on the verge of discussing specific research designs. (Our consideration of research designs begins in Chapter 10.) Because statistics and data analysis are integral components of the various research designs, a brief statistical refresher appears to be in order. We turn to this topic in Chapter 9.

Using Statistics to Answer Questions

CHAPTER 9

Descriptive Statistics

- Scales of Measurement • Measures of Central Tendency • Graphing Your Results • Calculating and Computing Statistics • Measures of Variability

Correlation

- The Pearson Product–Moment Correlation Coefficient

Inferential Statistics

- What Is Significant? • The *t* Test
- One-Tail Versus Two-Tail Tests of Significance
- The Logic of Significance Testing
- When Statistics Go Astray: Type I and Type II Errors

Effect Size

Just as detectives seek out clues and leads and gather data to help solve a case, psychologists, too, gather data to help answer research questions. After they have gathered the evidence, detectives must determine whether it is real (meaningful). Likewise, in later chapters, we will examine several statistical methods used to determine whether the results of an experiment are meaningful (significant). As we have seen, the term *significant* is used to describe those instances in which the statistical results are likely to have been caused by our manipulation of the independent variable (IV).

To understand better the nature of statistical significance, a closer look at statistics is in order. **Statistics** is a branch of mathematics that involves the collection, analysis, and interpretation of data. Researchers use various statistical techniques to aid them in several ways during the decision-making processes that arise when conducting research.

The two main branches of statistics assist your decisions in different ways. **Descriptive statistics** summarize any set of numbers so you can understand and talk about them more intelligibly. Researchers use **inferential statistics** to analyze data after they have conducted an experiment to determine whether the IV had a significant effect. Although we assume that you already have some familiarity with statistics, we have included several relevant formulas in Appendix B. We encourage you to review them at this time and as needed.

Statistics The branch of mathematics that involves the collection, analysis, and interpretation of data.

Descriptive statistics Procedures used to summarize a set of data.

Inferential statistics Procedures used to analyze data after an experiment is completed in order to determine whether the independent variable has a significant effect.

Descriptive Statistics

We use descriptive statistics when we want to summarize a set or distribution of numbers in order to communicate their essential characteristics. One of these essential characteristics is a measure of the typical or representative score, called a *measure of central tendency*. A

second essential characteristic that we need to know about a distribution is how much *variability* or spread exists in the scores. Before we discuss these measures of central tendency and variability, however, we must examine the measurements on which they are based.

Scales of Measurement

Measurement The assignment of symbols to events according to a set of rules.

Scale of measurement A set of measurement rules.

Nominal scale A scale of measurement in which events are assigned to categories.

Ordinal scale A scale of measurement that permits events to be rank ordered.

Interval scale A scale of measurement that permits rank ordering of events with the assumption of equal intervals between adjacent events.

We can define **measurement** as the assignment of symbols to events according to a set of rules. Your grade on a test is a symbol that stands for your performance; it was assigned according to a particular set of rules (the instructor's grading standards). The *particular* set of rules used in assigning a symbol to the event in question is known as a **scale of measurement**. The four scales of measurement that are of interest to psychologists are nominal, ordinal, interval, and ratio scales. How you choose to measure (i.e., which scale of measurement you use) the dependent variable (DV) directly determines the type of statistical test you can use to evaluate your data after you have completed your research project.

Nominal Scale The **nominal scale** is a simple classification system. For example, if you are categorizing the furniture in a classroom as tables or chairs, you are using a nominal scale of measurement. Likewise, recording responses to an item on a questionnaire as "agree," "undecided," or "disagree" reflects the use of a nominal scale of measurement. You assign the items being evaluated to mutually exclusive categories.

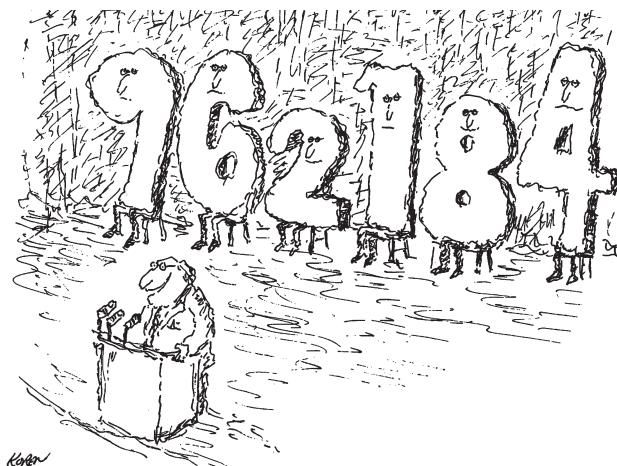
Ordinal Scale When you can rank order the events in question, you are using an **ordinal scale** of measurement. Notice that we indicated *only* that the events under consideration could be rank ordered; we did not indicate that the intervals separating the units were comparable. Although we can rank the winners in a track meet (i.e., first, second, third, fourth), this rank ordering does not tell us anything about how far apart the winners were. Perhaps it was almost a dead heat for first and second; maybe the winner was far ahead of the second-place finisher.

Interval Scale When you can rank order the events in question and equal intervals separate adjacent events, you are using an **interval scale**. For example, the temperatures on a Fahrenheit thermometer form an interval scale; rank ordering has been achieved *and* the difference between any two adjacent temperatures is the same, 1 degree. Notice that the interval scale does not have a true zero point, however. When you reach the "zero" point on a Fahrenheit thermometer, does temperature cease to exist? No, it's just very cold. Likewise, scores on tests such as the SAT and ACT are interval-scale measures.



Assume you are on a college admissions committee and you are reviewing applications. Each applicant's ACT score forms an integral part of your review. You have just come across an applicant who scored 0 on the verbal subtest. What does this score tell you?

The score of 0 should not be interpreted as meaning that this individual has absolutely no verbal ability. Because ACT scores are interval-scale measurements, there is no true zero. A score of 0 should be interpreted as meaning that the individual is very low in that ability. The same could be said for 0 scores on the wide variety of tests, questionnaires, and personality inventories routinely used by psychologists in personality research. The presence of a true zero is characteristic only of the ratio scale of measurement.



"Tonight, we're going to let the statistics speak for themselves."

© The New Yorker Collection 1974 Edward Koren from cartoonbank.com.
All Rights Reserved.

Ratio Scale The **ratio scale** of measurement takes the interval scale one step farther. Like the interval scale, the ratio scale permits the rank ordering of scores with the assumption of equal intervals between them, *but it also assumes the presence of a true zero point*. Physical measurements, such as the amplitude or intensity of sound or light, are ratio measurements. These measurements can be rank ordered, and there are equal intervals between adjacent scores. However, when a sensitive measuring device reads 0, there is nothing there. Because of the true zero point, the ratio scale allows you to make ratio comparisons, such as “twice as much” or “half as much.”

Our discussion of scales of measurement has progressed from the nominal scale, which provides the least amount of information, to the ratio scale, which provides the greatest amount of information. When psychologists evaluate changes in the DV, they try to use a scale of measurement that will provide the most information; frequently, they select interval scales because they do not use measurements that have a true zero.

We now turn to the topic of central tendency. Keep in mind that the scales of measurement directly determine which measure of central tendency you will use.

Ratio scale A scale of measurement that permits rank ordering of events with the assumptions of equal intervals between adjacent events and a true zero point.

Measures of Central Tendency

Measures of central tendency, such as the mode, median, and mean, tell us about the typical score in a distribution.

Mode The score in a distribution that occurs most often.

Mode The **mode** is the number or event that occurs most frequently in a distribution. If students reported the following work hours per week

12, 15, 20, 20, 20

the mode would be 20.

$$\text{Mode} = 20$$

Although the mode can be calculated for any scale of measurement, it is the only measure of central tendency that can be used for nominal data.

Median The number that divides a distribution in half.

Median The **median** is the number or score that divides the distribution into equal halves. To be able to calculate the median, you must first rank order the scores. Thus, if you started with the following scores

56, 15, 12, 20, 17

you would have to rank order them as follows:

12, 15, 17, 20, 56

Now it's an easy task to determine that 17 is the median (Mdn):

$$\text{Mdn} = 17$$

What if you have an even number of scores, as in the following distribution?

1, 2, 3, 4, 5, 6

In this case the median lies halfway between the two middle scores (3 and 4). Thus, the median would be 3.5, halfway between 3 and 4. The median can be calculated for ordinal, interval, and ratio data.

Mean The arithmetic average of a set of numbers; found by adding all the scores in a set and then dividing by the number of scores.

Mean The **mean** is defined as the arithmetic average. To find the mean we add all the scores in the distribution and then divide by the number of scores we added. For example, assume we start with

12, 15, 18, 19, 16

We use the Greek letter sigma, Σ , to indicate the sum. If X stands for the numbers in our distribution, then ΣX means to add the numbers in our distribution. Thus, $\Sigma X = 80$. If N stands for the number of scores in the distribution, then the mean would equal $\Sigma X/N$. For the previous example, $80/5 = 16$. The sum of these numbers is 80, and the mean is 16 ($80/5$). The mean is symbolized by M .

You may recall from your statistics class that \bar{X} stood for the mean. We haven't arbitrarily changed symbols on you. Because M stands for the mean in APA-format papers (see Chapter 14), we chose to use it instead of \bar{X} . Thus, $M = 16$. You can calculate the mean for interval and ratio data, but not for nominal and ordinal data.

Choosing a Measure of Central Tendency Which measure of central tendency should you choose? The answer to that question depends on the type of information you are seeking and the scale of measurement you are using. If you want to know which score occurred most often, then the mode is the choice. However, the mode may not be very representative of the other scores in your distribution. Consider the following distribution:

1, 2, 3, 4, 5, 11, 11

In this case the mode is 11. Because all the other scores are considerably smaller, the mode does not accurately describe the typical score.

The median may be a better choice to serve as the representative score because it takes into account all the data in the distribution; however, there are drawbacks with this choice. The median treats all scores alike; differences in magnitude are not taken into account. Thus, the median for *both* of the following distributions is 14:

$$\text{Distribution 1: } 11, 12, 13, \mathbf{14}, 15, 16, 17 \quad \text{Mdn} = 14$$

$$\text{Distribution 2: } 7, 8, 9, \mathbf{14}, 23, 24, 25 \quad \text{Mdn} = 14$$

When we calculate the mean, however, the *value* of each number is taken into account. Although the medians for the two distributions above are the same, the means are not:

$$\text{Distribution 1: } 11, 12, 13, 14, 15, 16, 17$$

$$\Sigma X = 98 \quad M = 98/7 \quad M = 14$$

$$\text{Distribution 2: } 7, 8, 9, 14, 23, 24, 25$$

$$\Sigma X = 110 \quad M = 110/7 \quad M = 15.71$$

The fact that the mean of Distribution 2 is larger than that of Distribution 1 indicates that the value of each individual score has been taken into account.

Because the mean takes the value of each score into account, it usually provides a more accurate picture of the typical score, and it is the measure of central tendency favored by psychologists. On the other hand, there are instances in which the mean may be misleading. Consider the following distribution of charitable donations:

Charitable donations: \$1, \$1, \$1, \$5, \$10, \$10, \$100

Mode = \$1

Mdn = \$5

Mean = \$128/7 $M = \$18.29$

If you wanted to report the “typical” gift, would it be the mode? Probably not. Even though \$1 is the most frequent donation, this amount is *substantially* smaller than the other donations, and more people made contributions over \$1 than made the \$1 contribution. What about the median? You see that \$5 appears to be more representative of the typical donation; there are an equal number of higher and lower donations. Would the mean be better? In this example the mean is substantially inflated by one large donation (\$100); the mean is \$18.29 even though six of the seven donations are \$10 or under. Although reporting the mean in this case may look good on a report of giving, it does not reflect the typical donation.

The lesson to be learned from this example is that when you have only a limited number of scores in your distribution, the mean may be inflated (or deflated) by extremely large (or extremely small) scores. The median may be a better choice as your measure of central tendency in such instances. As the number of scores in your distribution increases, the influence of extremely large (or extremely small) scores on the mean decreases. Look what happens if we collect two additional \$5 donations:

Charitable donations: \$1, \$1, \$1, \$5, \$5, \$5, \$10, \$10, \$100

Mode = \$1 and \$5

Mdn = \$5

Mean = \$138/9 $M = \$15.33$

Note we now have two values (\$1 and \$5) for the mode (i.e., a *bimodal distribution*). The median stays the same (\$5); however, the mean has decreased from \$18.29 to \$15.33; the addition of only two more low values moved the mean closer to the median.

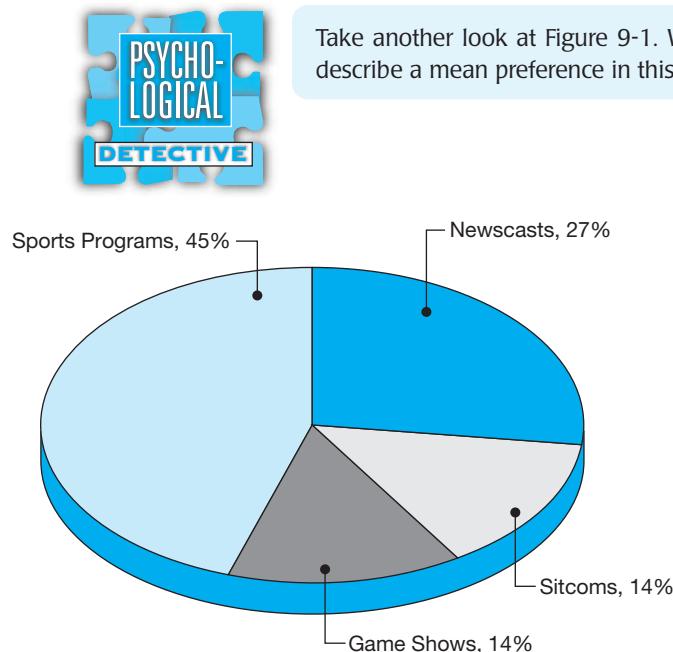
Graphing Your Results

After you have calculated a measure of central tendency, you can convey this information to others. If you have only one set of scores, the task is simple: You write down the value as part of your paper or report.

What if you are dealing with several groups or sets of numbers? Now the task is complicated, and the inclusion of several numbers in a paragraph of text might be confusing. In such cases a graph or figure can be used to your advantage; a picture may well be worth a thousand words. It is not uncommon to see a detective use a chart or graph to help make a point in the solution of a case. In preparing a research report, psychologists also use graphs effectively. There are several types of graphs for the researcher to choose from. Your choice of graphs will be determined by which one depicts your results most effectively *and* by the scale of measurement you used. For example, if you used a nominal scale of measurement, then you would probably use a pie chart, a histogram, a bar graph, or a frequency polygon.

Pie chart Graphical representation of the percentage allocated to each alternative as a slice of a circular pie.

Pie Chart If you are dealing with percentages that total 100%, the familiar pie chart may be a good choice. The **pie chart** depicts the percentage represented by each alternative as a slice of a circular pie. The larger the slice, the greater the percentage. For example, if you surveyed college men to determine their TV viewing preferences, you might display your results as a pie chart (see Figure 9-1). From the hypothetical data presented in Figure 9-1 we can see that the mode is sports programs.



Take another look at Figure 9-1. Why would it not be appropriate to describe a mean preference in this instance?

FIGURE 9-1 TV Viewing Preferences of College Men.

To answer this question, ask yourself what scores we would use to obtain the mean. We know there are four categories of TV preference, and we know the percentage preferring each category. We could add the percentages for each category and then divide by the number of categories. The resulting number would tell us that there is an average of 25% per category; unfortunately, that number does not tell us much about the “mean preference.” We would have to have individual scores, in the form of interval or ratio data, before we could calculate a mean preference. These data are simply not available.

Histogram We can use a **histogram** to present our data in terms of frequencies per category. When we study a *quantitative variable*, we construct a histogram. Quantitative categories are ones that can be numerically ordered. The levels or categories of a quantitative variable must be arranged in a numerical order. For example, we may choose to arrange our categories from smallest to largest, or vice versa. Figure 9-2 shows a histogram for the age categories of participants in a developmental psychology research project.

Histogram A graph in which the frequency for each category of a quantitative variable is represented as a vertical column that touches the adjacent column.

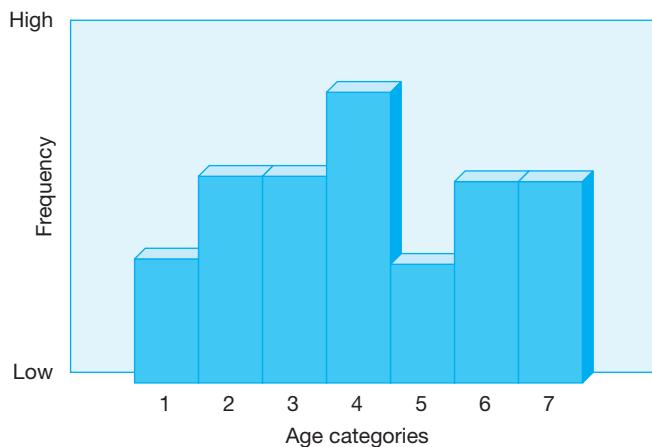


FIGURE 9-2 Histogram Depicting the Frequency of Participants in Various Age Categories in a Developmental Psychology Research Project. Note that the sides of adjacent columns touch.

Bar Graph The **bar graph** also presents data in terms of frequencies per category; however, we are using *qualitative categories* when we construct a bar graph. Qualitative categories are ones that cannot be numerically ordered. For example, single, married, divorced, and remarried are qualitative categories; there is no way to order them numerically.

Figure 9-3 shows a bar graph for the sports and fitness preferences of women and girls who are frequent participants in such activities. Placing a space between the bars lets the reader know that qualitative categories are being reported. You can see at a glance that the number per category and the type of activities differ dramatically between the two groups. Think of how many words it would take to write about these results rather than present them in graph form.

Bar graph A graph in which the frequency for each category of a qualitative variable is represented as a vertical column. The columns of a bar graph do not touch.

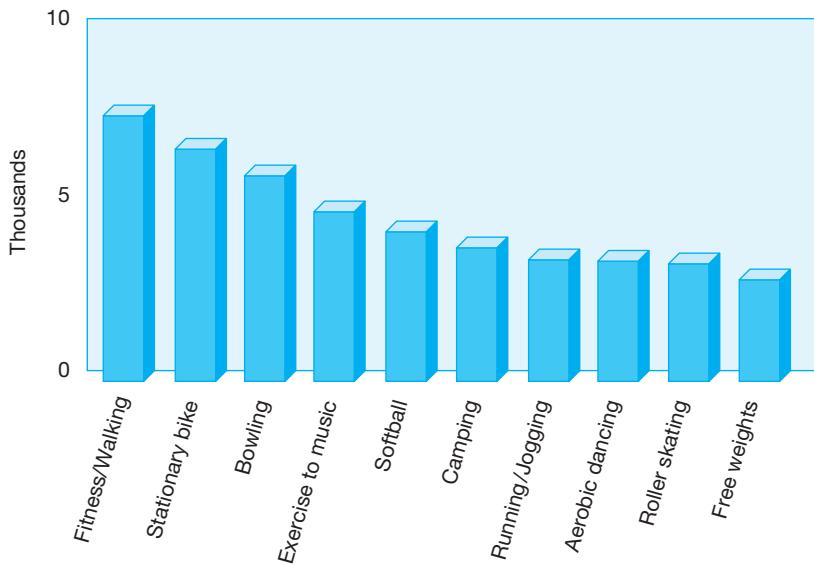


FIGURE 9-3 Bar Graphs Depicting the Sports and Fitness Preferences of Women and Girls Who Are Frequent Participants in Such Activities. Because a bar graph depicts a qualitative variable, the bars do not touch.

Source: *The American Enterprise*, a national magazine of politics, business, and culture, September/October 1993, p. 101. TAEmag.com.

Frequency polygon A graph that is constructed by placing a dot in the center of each bar of a histogram and then connecting the dots.

Line graph A graph that is frequently used to depict the results of an experiment.

Ordinate The vertical or y axis of a graph.

Abscissa The horizontal or x axis of a graph.

Frequency Polygon If we mark the middle of the crosspiece of each bar in a histogram (see Figure 9-4A) with a dot, connect the dots, and remove the bars, we have constructed a **frequency polygon** (see Figure 9-4B).

The frequency polygon, like the histogram, displays the frequency of each number or score. The only difference between the two is that we used bars in the histogram and connected dots in the frequency polygon.

Line Graph Researchers frequently present the results of psychological experiments as a **line graph**. In constructing a line graph, we start with two axes or dimensions. The vertical or y axis is known as the **ordinate**; the horizontal or x axis is known as the **abscissa** (see Figure 9-5). We plot our scores or data (the DV) on the ordinate. The values of the variable we manipulated (the IV) are plotted on the abscissa.

How tall should the y axis be? How long should the x axis be? A good rule of thumb is for the y axis to be approximately two-thirds as tall as the x axis is long (see Figures 9-5 and 9-6). Other configurations will give a distorted picture of the data. For example, if the ordinate is considerably shorter, differences between groups or treatments will be obscured (see Figure 9-7A, page 180), whereas lengthening the ordinate tends to exaggerate differences (see Figure 9-7B).

In Figure 9-6 we have plotted the results of a hypothetical experiment that evaluated the effects of different levels of stress on making correct landing decisions by air traffic controllers. As you can see, as stress increased, the number of correct responses increased. What if we had tested two different groups of participants,

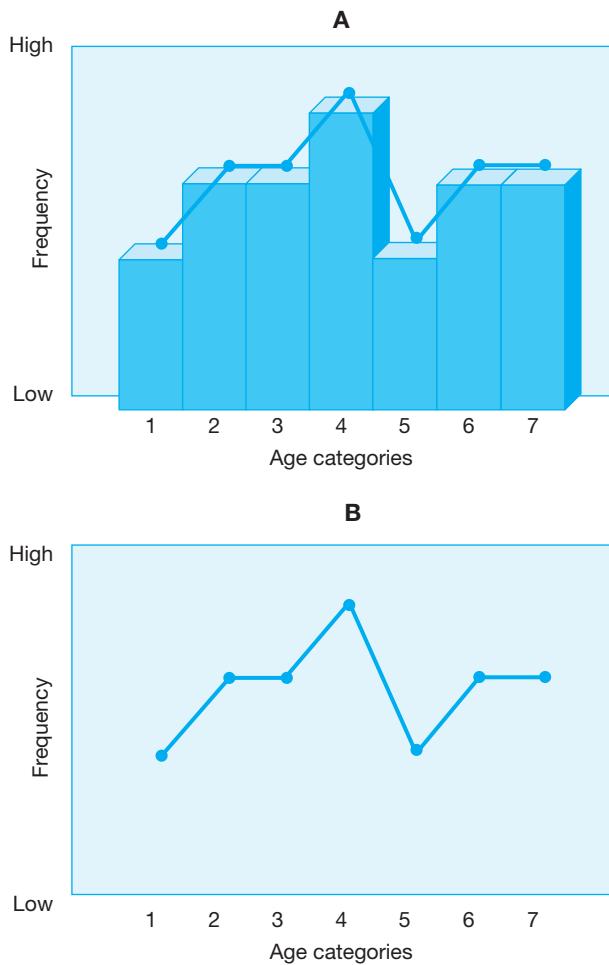


FIGURE 9-4 The Frequency Polygon Is Constructed by Placing a Dot in the Center of Each Bar of a Histogram and Connecting the Dots (A) and Removing the Bars (B). The frequency polygon, like the histogram, displays the frequency of each score or number.

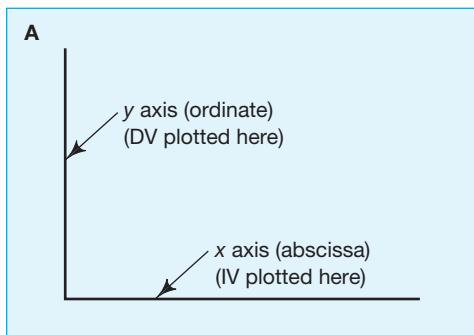


FIGURE 9-5 The Ordinate or y Axis and Abscissa or x Axis. The ordinate should be about $\frac{2}{3}$ the size of the abscissa to portray the data as clearly as possible.

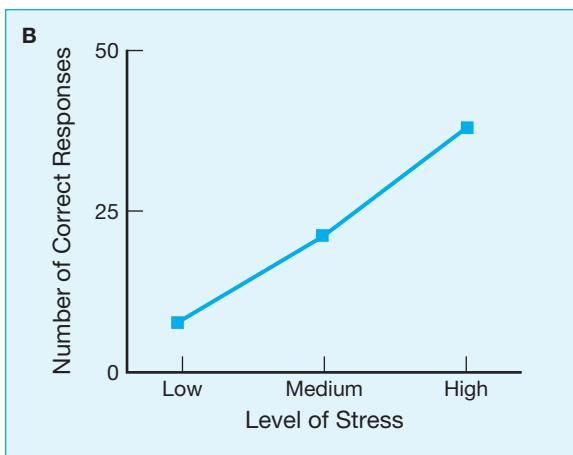


FIGURE 9-6 Results of a Hypothetical Experiment Investigating the Effects of Stress on Correct Response by Air Traffic Controllers.

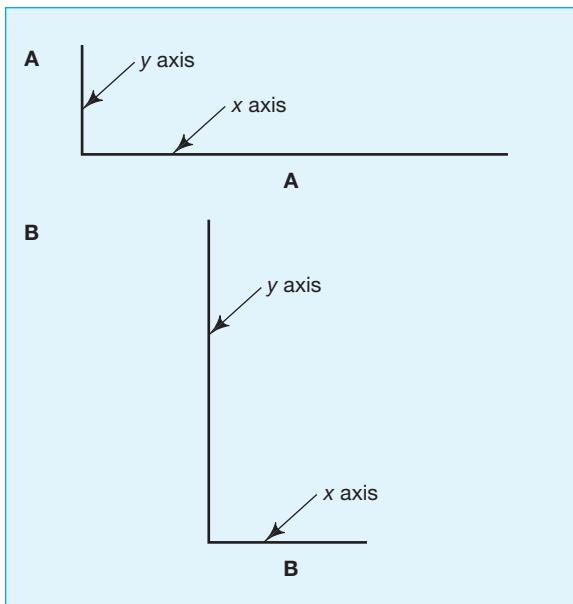


FIGURE 9-7 Altering the *x* (Abscissa) or *y* (Ordinate) Axis Can Distort the Results of an Experiment. **A.** If the ordinate is considerably shorter than the abscissa, significant effects can be obscured. **B.** If the ordinate is considerably longer than the abscissa, very small effects can be exaggerated.

college students and air traffic controllers? How would we display the results of both groups on the same graph? No problem. All we must do is add the data points for the second group and a legend or box that identifies the groups (see Figure 9-8). Now, we can see at a glance that the air traffic controllers, whose occupation is very stressful, made more correct responses as stress levels increased; the converse was true for the college students.

When you graph the results of an experiment in which more than one variable is used, how do you know which IV to plot on the abscissa? Although there is no fixed rule, a good guideline is to plot the variable having the greater number of levels on the abscissa. Thus, in Figure 9-8 the three levels of stress were plotted on the abscissa, rather than the two levels of participants.

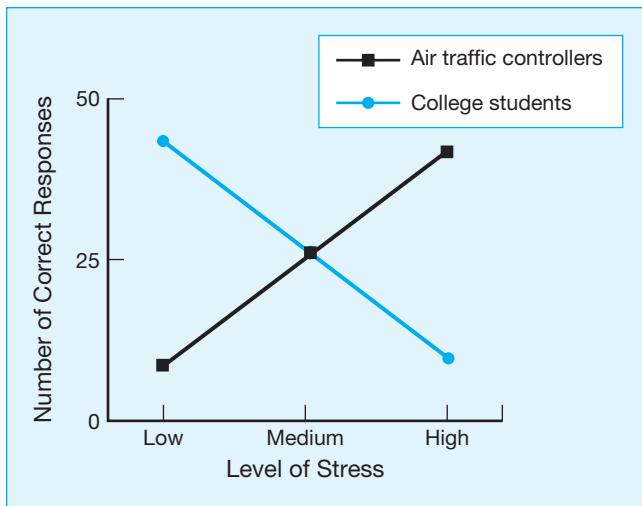


FIGURE 9-8 Results of a Hypothetical Experiment Investigating the Effects of Stress on Correct Response by Air Traffic Controllers and College Students.



Why would you choose to plot the IV with the greatest number of levels on the abscissa?

By plotting the variable with the greatest number of levels on the abscissa, you reduce the number of lines that will appear on your graph. The fewer lines, the less difficulty the reader will have in interpreting your graph. For example, had we plotted the type of participants in our stress experiment on the abscissa, then Figure 9-8 would have had three lines, one for each level of stress. We will discuss the accepted APA format for preparing graphs and tables in Chapter 14.

Calculating and Computing Statistics

Remember, this is *not* a statistics text—we assume you have already taken a statistics course; therefore, we will not review formulas for all the various statistical calculations and tests you will encounter in this book. You can find those formulas in Appendix B. Use them as needed. Your calculating skills may be a little rusty, but all statistical formulas merely require addition, subtraction, multiplication, division, and finding square roots—not all that challenging for a college student . . . especially one with a calculator.

By the same token, most psychologists (and probably most psychology students) rarely use hand computation techniques for statistics after their initial statistics course; the vast majority use a computer package of some sort to analyze data they collect. Of course, these computer packages vary widely. You may have access to a large and powerful statistics package owned by your school or department (some standard packages are SPSS, SAS, and BMD; you probably have Microsoft Excel® on your computer). Alternatively, you may have access to

a smaller statistics program; some schools even require students to buy a statistics software program when they take the statistics course. In any case, you are likely to have access to a computerized statistical analysis program. We cannot begin to give instructions about how to operate the particular program you might have access to—there are simply too many programs. Throughout the chapters that deal with statistics, we will attempt to give you some *general* hints about how to interpret the output you receive from such programs.

Measures of Variability

Although measures of central tendency and graphs convey considerable information, there is still more we can learn about the numbers we have gathered. We also have to know about the variability in our data.

Imagine that our instructor just returned your last psychology test; your score is 64. What does that number tell you? By itself it may not mean very much. You ask your professor for additional information and find that the class mean was 56. You feel better because you were

above the mean; however, after a few moments of reflection you realize you need still more information. How were the other scores grouped? Were they all clustered close to the mean or did they spread out considerably? The amount of **variability** or spread in the other scores will have a bearing on the standing of your score. If most of the other scores are very close to the mean, then your score will be among the highest in the class. If the other scores are spread out widely around the mean, then your score will not be one of the strongest. Obviously, you need a measure of variability to provide a complete picture of these data. Range and standard deviation are two measures of variability frequently reported by psychologists.

Range The **range** is the easiest measure of variability to calculate; you rank order the scores in your distribution and then subtract the smallest score from the largest to find the range. Consider the following distribution:

$$1, 1, 1, 1, 5, 6, 6, 8, 25$$

When we subtract 1 (the smallest score) from 25 (the largest score), we find that the range is 24:

$$\text{Range: } 25 - 1 = 24$$

However, other than telling us the difference between the largest and smallest scores, the range does not provide much information. Knowing the range is 24 does not tell us about the distribution of the scores we just considered. Consider Figure 9-9.

The range is the same in Parts A and B of Figure 9-9; however, the spread of the scores differs drastically between these two distributions. Most of the scores cluster in the center of the first distribution (Figure 9-9A), whereas the scores are spread out more evenly in the second distribution (Figure 9-9B). We must turn to another measure, the standard deviation, to provide additional information about how the scores are distributed.

Variance A single number that represents the total amount of variation in a distribution; also the square of the standard deviation, σ^2 .

Variance and Standard Deviation To obtain the standard deviation, we must first calculate the variance. You can think of the **variance** as a single number that represents the total amount of variability in the distribution. The larger the number, the greater the total spread of the scores. The variance and standard deviation are based on how much each score in the distribution deviates from the mean.

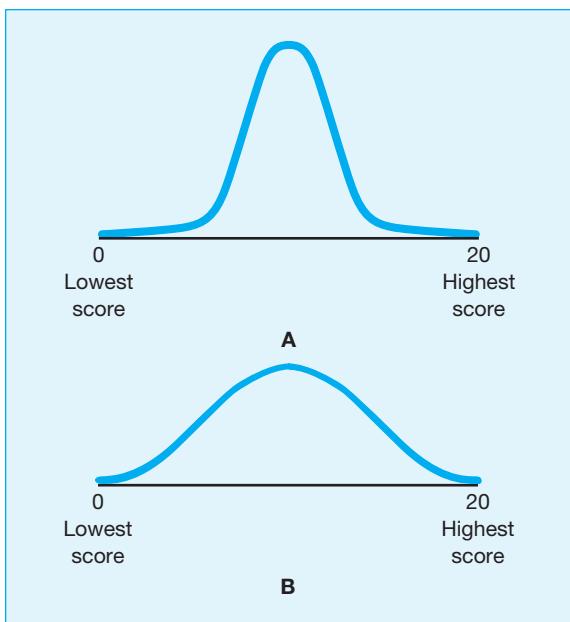


FIGURE 9-9 The Range Does Not Provide Much Information About the Distribution Under Consideration. Even though the range is the same, these two distributions differ drastically.

When researchers conduct experiments, they use a sample of participants to provide information (an estimate) about an entire population. As an example we calculated the variance of the set of nine numbers for which we computed the range and found it to be 58.25. (See Appendix B for the formula to calculate the variance.) Once we have the variance, we can use it to find the standard deviation.

Interpreting the Standard Deviation To find the **standard deviation (SD)**, all we have to do is take the square root of the variance. Using our variance of 58.25,

$$\begin{aligned} SD &= \sqrt{\text{variance}} \\ &= \sqrt{58.25} \\ &= 7.63 \end{aligned}$$

Standard deviation

Square root of the variance; has important relations to the normal curve.

As with the variance, the larger the standard deviation is, the greater the variability or spread of scores will be.

A sample computer printout listing the mean, variance, and standard deviation for the nine scores we used when we calculated the range appears in Table 9-1. Rather than providing printouts from a specific statistics software package, we are using generic printouts in the text. Statistics packages provide slightly different information; we show you the information you might reasonably expect to find on your printout. As you can see, the computer informs us that we entered nine numbers. The mean of these nine numbers is 6.00, the variance is 58.25, and the standard deviation is 7.63.

TABLE 9-1 Computer Printout Showing Mean, Standard Deviation, and Variance

Mean	SD	Variance	Range	N
6.00	7.63	58.25	24.00	9

Normal distribution A symmetrical, bell-shaped distribution having half the scores above the mean and half the scores below the mean.

Now that we have found the standard deviation, what does it tell us? To answer that question we must consider the normal distribution (also called the *normal curve*). The concept of the **normal distribution** is based on the finding that as we increase the number of scores in our sample, many distributions of interest to psychologists become symmetrical or bell shaped. (Sometimes the normal distribution is called the *bell curve*.) The majority of the scores cluster around the measure of central tendency, with fewer and fewer scores occurring as we move away from it. As you can see from Figure 9-10, the mean, median, and mode of a normal distribution all have the same value.

Normal distributions and standard deviations are related in interesting ways. For example, distances from the mean of a normal distribution can be measured in standard deviation units (SD). Consider a distribution with an M of 56 and an SD of 4; a score of 60 falls 1 SD above the mean (+1 SD), whereas a score of 48 is 2 SD below the mean (-2 SD), and so on. As you can see from Figure 9-11, 34.13% of all the scores in *all* normal distributions occur between the mean and 1 SD *above* the mean.

Likewise, 34.13% of all the scores in a distribution occur between the mean and 1 SD *below* the mean. Another 13.59% of the scores occur between 1 and 2 SD *above* the mean; another 13.59% of the scores occur between 1 and 2 SD *below* the mean. Thus, slightly over 95% of all the scores in a normal distribution occur between 2 SD *below* the mean and 2 SD

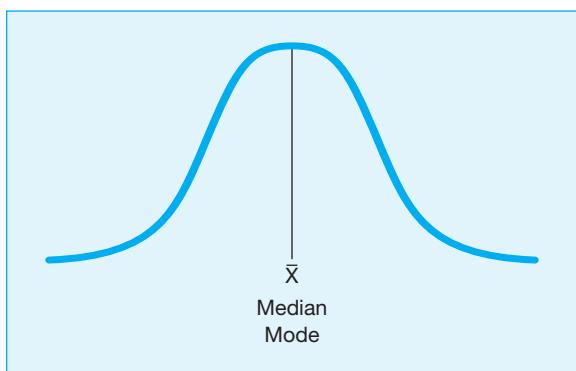


FIGURE 9-10 A Symmetrical or Bell-Shaped Normal Distribution. Note that the mean, median, and mode coincide in a normal distribution.

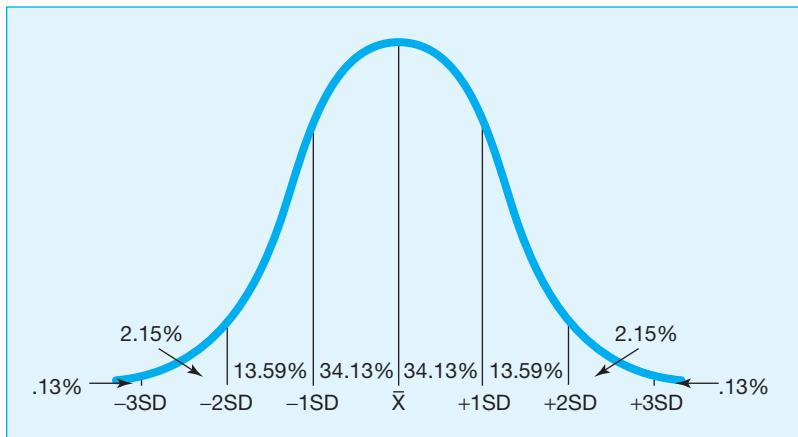


FIGURE 9-11 The Relation of Standard Deviations and the Normal Distribution.

above the mean. Exactly 2.28% of the scores occur *beyond 2 SD above* the mean; another 2.28% of the scores occur *beyond 2 SD below* the mean. It is important to remember that these percentages hold true for *all* normal distributions.



Review Figure 9-11 for a moment. Why isn't the percentage of scores from 0 to 1 *SD* above (or below) the mean the same as the percentage of scores from 1 to 2 *SD* above (or below) the mean?

As we move away from the mean (either above or below), the scores become progressively different from the mean. Because larger scores occur less frequently and the scores between 1 and 2 *SD* are larger than those from 0 to 1 *SD*, the percentage of scores from 1 to 2 *SD* will be lower than the percentage of scores from 0 to 1 *SD*.

Now, let's return to your test score of 64. You know the mean of the class is 56. If the instructor also tells you that the $SD = 4$, what would your reaction be? Your score of 64 would be 2 *SD* above the mean; you should feel pretty good. Your score of 64 puts you in the top 2.28% of the class (100% minus 50% of the scores below the mean and minus 34.13% from the mean to 1 *SD* above the mean and minus 13.59% that occurs between 1 and 2 *SD* above the mean). (See Figure 9-12A.)

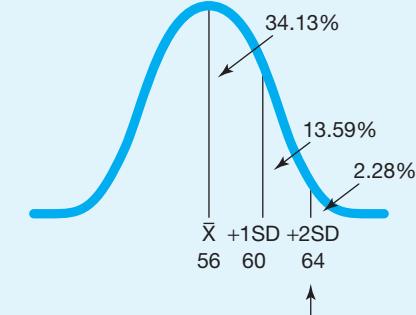
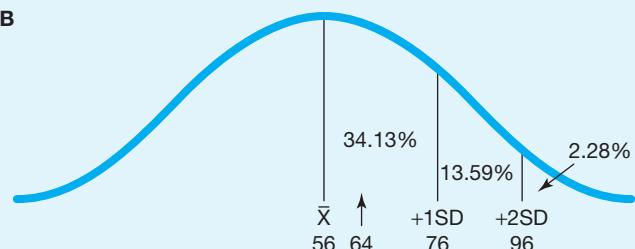
A**B**

FIGURE 9-12 **A.** A score of 64 is exceptionally good when the mean is 56 and the SD is 4. **B.** The same score is not as highly regarded when the SD is 20.

What if your instructor told you the *SD* was 20? Now your score of 64 does not stand up as well as it did when the *SD* was 4. You are above the mean but a long way from being even 1 *SD* above the mean (see Figure 9-12B).

Because the percentage of the scores that occurs from the mean to the various *SD* units is the same for *all* normal distributions, we can compare scores from different distributions by discussing them in terms of standard deviations above or below the mean. Consider the following scores:

Test #	Your Score	M	SD	Relation of Your Score to the M
1	46	41	5	1 <i>SD</i> above
2	72	63	4	Over 2 <i>SD</i> above
3	93	71	15	Over 1 <i>SD</i> above

Even though your scores, the means, and the standard deviation values differ considerably, we can determine how many *SD* units away from the mean each of your scores is. In turn, we can compare these differences. When these comparisons are made, we find that your scores are consistently 1 *SD* or more above the mean. Thus, you are consistently in at least the top 15.87% of the class (100% minus 50% of the scores below the mean and minus 34.13% of the scores from the mean to 1 *SD* above the mean). By comparing scores from various distributions in this manner, we are able to see patterns and suggest what might occur in the future. Another type of descriptive statistic, the correlation coefficient, is also used for predictive purposes. We turn to this topic next.



Courtesy of Sidney Harris.

Unlike the warning on this truck, psychologists view data and statistical procedures as tools to help answer research questions.

■ REVIEW SUMMARY

1. **Statistics** involves the collection, analysis, and interpretation of data.
2. **Measurement** is the assignment of symbols to events according to a set of rules. A **scale of measurement** is a particular set of measurement rules.
3. A **nominal scale** is a simple classification system, whereas events can be rank ordered when an **ordinal scale** is used. Equal intervals separate rank-ordered events in an **interval scale**. The addition of a true zero to an interval scale results in a **ratio scale**.
4. **Descriptive statistics**, which summarize sets of numbers, include measures of central tendency and variability.
5. The **mode** is the most frequent score, whereas the **median** divides a distribution into two equal halves. The **mean** is the arithmetic average. Depending on the nature of the distribution, these measures of central tendency may not reflect the typical score equally well. They are, however, identical in a normal distribution.
6. Graphs, such as the **pie chart**, **bar graph**, **histogram**, and **frequency polygon**, are often used to depict frequencies or percentages.
7. The **line graph** is used to depict experimental results. The DV is plotted on the vertical (y) axis (**ordinate**), and the IV is plotted on the horizontal (x) axis (**abscissa**). A 2:3 relation of y to x axes produces a representative figure.
8. Measures of **variability** include the **range** (difference between high and low scores) and **standard deviation** (SD ; square root of the variance). The **variance** is a single number that represents the total amount of variability that is present.
9. The standard deviation conveys considerable information about the **normal distribution** that is under consideration.

■ Check Your Progress

1. Matching

1. inferential statistics	A. assignment of symbols to events
2. descriptive statistics	B. rank order
3. measurement	C. putting events into categories
4. nominal scale	D. equal intervals plus a true zero
5. ordinal scale	E. used to summarize a set of numbers
6. interval scale	F. equal intervals
7. ratio scale	G. used to analyze data after an experiment
2. The number that occurs most frequently is the

a. mean	b. median	c. mode	d. harmonic
---------	-----------	---------	-------------
3. When you are dealing with a normal distribution of scores, which measure of central tendency is preferred? Why?
4. A _____ presents data in terms of frequencies per category.

a. pie chart	b. line graph	c. bimodal distribution	d. histogram
--------------	---------------	-------------------------	--------------

5. You are constructing a line graph to depict the results of an experiment you just completed. What is the ordinate? What is the abscissa? What will be plotted on each of them?
6. Why does the range not convey much information about variability?
7. The _____ is a single number that represents the total amount of variability in a distribution.
 - a. variance
 - b. standard deviation
 - c. range
 - d. mean
8. How does the standard deviation relate to the normal curve?

Correlation

Just as it does in the successful completion of a detective case, prediction plays an important role in psychology. Nowhere is this aspect of psychology more apparent than when moving from high school to college. You probably took a college entrance examination while you were in high school. Based on the results of this exam, a prediction about your grades in college was made. Similarly, should you plan to go on for graduate training after you complete your undergraduate degree, you probably will take another entrance examination. Depending upon your area of interest, you might take the Graduate Record Examination (GRE), the Law School Admission Test (LSAT), the Medical College Admission Test (MCAT), or some other similar test.

Such predictions are based on the correlation coefficient. Sir Francis Galton (1822–1911) developed the basic ideas of correlation. Galton, who was independently wealthy, devoted his time to studying and investigating those things that interested him. According to E. G. Boring (1950), the eminent historian of psychology, Galton “was a free-lance and a gentleman scientist. He was forever seeing new relationships and working them out, either on paper or in practice. No field was beyond his possible interest, no domain was marked off in advance as being out of his province” (p. 461). For example, Galton studied such varied topics as the weather and fingerprints. He also proposed that a person’s intelligence was directly related to the quality of the nervous system: The better the nervous system, the higher the intelligence. To be able to measure the predictive relation between these two variables, Galton’s assistant, Karl Pearson (1857–1936), developed the correlation coefficient.

A **correlation coefficient** is a single number that represents the degree of relation (i.e., “co-relation”) between two variables. The value of a correlation coefficient can range from -1 to $+1$.

A correlation coefficient of -1 indicates that there is a *perfect negative relation* (see Figure 9-13) between the two variables of interest. That is, whenever we see an increase of 1 unit in one variable, there is always a proportional decrease in the other variable.

Consider the following scores on Tests X and Y:

	Test X	Test Y
Student 1	49	63
Student 2	50	61
Student 3	51	59
Student 4	52	57
Student 5	53	55

Correlation coefficient

A single number representing the degree of relation between two variables.

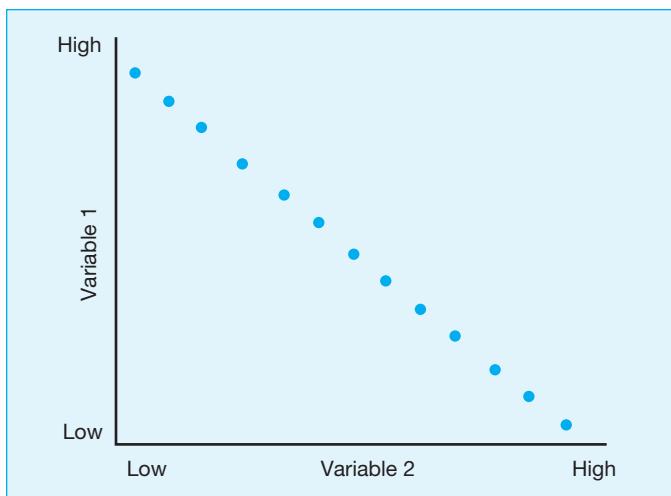


FIGURE 9-13 A Perfect Negative Correlation.

For each unit of increase in a score on Test X, there is a corresponding *decrease* of 2 units in the score on Test Y. Given this information, you are able to predict that if Student 6 scores 54 on Test X, that student's score on Test Y will be 53.

As you saw in Chapter 4, a *zero correlation* means that there is *little or no relation* between the two variables (see Figure 9-14). As scores on one variable increase, scores on the other variable may increase, decrease, or be the same. Hence, we are not able to predict how you will do on Test Y by knowing your score on Test X. A correlation coefficient does not have to

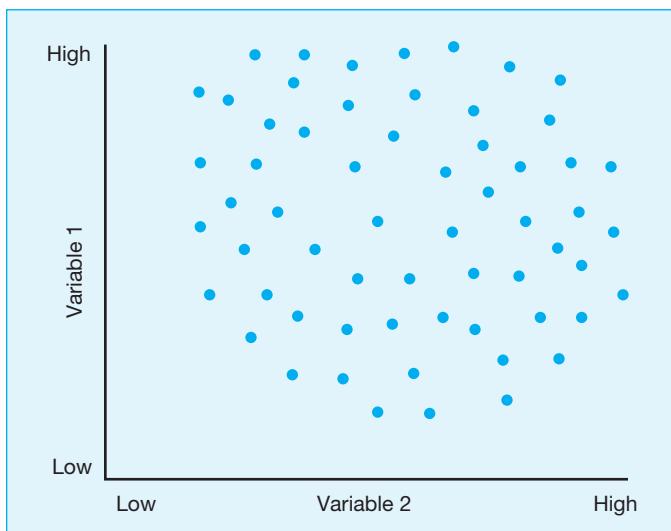


FIGURE 9-14 A Zero Correlation.

be exactly 0 to be considered a zero correlation. The inability to make good predictions is the key consideration. Two sets of scores having a near-zero correlation might look like this:

	Test X	Test Y
Student 1	58	28
Student 2	59	97
Student 3	60	63
Student 4	61	60
Student 5	62	50

In this case the correlation between Test X and Test Y is 0.04. A correlation that small indicates that you will not be able to predict Test Y scores by knowing Test X scores; you are dealing with a zero correlation or no relation.

A correlation coefficient of +1 indicates that there is a *perfect positive relation* between the two sets of scores (see Figure 9-15). That is, when we see an increase of 1 unit in one variable, we always see a proportional increase in the other variable. Consider the following scores on Tests X and Y:

	Test X	Test Y
Student 1	25	40
Student 2	26	43
Student 3	27	46
Student 4	28	49
Student 5	29	52

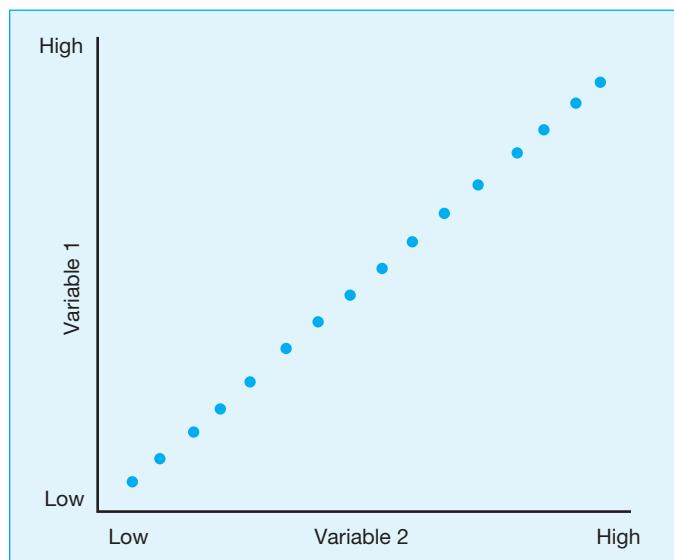


FIGURE 9-15 A Perfect Positive Correlation.

In this example, there is an *increase* of 3 units in the score on Test Y for every unit increase on Test X. The perfect positive correlation leads you to predict that if Student 6 scores 30 on Test X, then his or her score on Test Y will be 55.



Now that we have reviewed the basics of correlation, we would like you to think about the following question. Do perfect (either positive or negative) correlations occur frequently in the real world? Why or why not?

The existence of a perfect correlation indicates that there are no other factors present that influence the relation we are measuring. This situation rarely occurs in real life. Think about correlating test scores. To obtain a perfect correlation, all of the participants would have to study and learn the same amount for each test. This situation is not likely to occur. Such factors as fatigue, illness, boredom, and distractions will likely have an effect and make the correlation less than perfect.

The Pearson Product–Moment Correlation Coefficient

The most common measure of correlation is the Pearson product–moment correlation coefficient (r), which was developed by Galton's assistant Karl Pearson. This type of correlation coefficient is calculated when both the X variable and the Y variable are interval or ratio scale measurements and the data appear to be linear. Other correlation coefficients can be calculated when one or both of the variables are not interval or ratio scale measurements or when the data do not fall on a straight line.

Examples of computer printouts for perfect positive and perfect negative correlations appear in Table 9-2. As you can see, the correlation of Tests X and Y with themselves is always

TABLE 9-2 Computer Printout for (A) Perfect Negative Correlation and (B) Perfect Positive Correlation

A. Perfect Negative Correlation

PEARSON CORRELATION MATRIX

	TEST X	TEST Y
TEST X	1.00	
TEST Y	-1.00	1.00

NUMBER OF OBSERVATIONS: 5

B. Perfect Positive Correlation

PEARSON CORRELATION MATRIX

	TEST X	TEST Y
TEST X	1.00	
TEST Y	1.00	1.00

NUMBER OF OBSERVATIONS: 5

1.00; however, the correlation of Test X with Test Y is -1.00 (Table 9-2A) when the relation is perfect negative and $+1.00$ (Table 9-2B) when the relation is perfect positive. Because some computers do not provide a probability for correlations, you may have to consult a table of correlation probabilities to determine whether a particular correlation is significant (see Table A-3 in Appendix A).

Review Figures 9-13, 9-14, and 9-15; they will help you visualize the various correlations we have discussed. Perfect positive and perfect negative correlations always fall on a straight line, whereas nonperfect correlations do not. You will find, however, that the more the scores cluster close together and form a straight line, the stronger (i.e., larger) the correlation coefficient will be. For positive correlations, the trend of the points is from lower left to upper right, whereas for negative correlations, the trend is from upper left to lower right. There is no consistent pattern for a zero correlation.

Although descriptive statistics can tell us a great deal about the data we have collected, they cannot tell us everything. For example, when we conduct an experiment, descriptive statistics cannot tell us whether the IV we manipulated had a significant effect on the behavior of the participants we tested or whether the results we obtained would have occurred by chance. To make such determinations we must conduct an inferential statistical test.

Inferential Statistics

After you have conducted an experiment, you perform a statistical test on the data that you have gathered. The results of this test will help you decide whether the IV was effective. In other words, we shall decide whether our statistical result is *significant*.

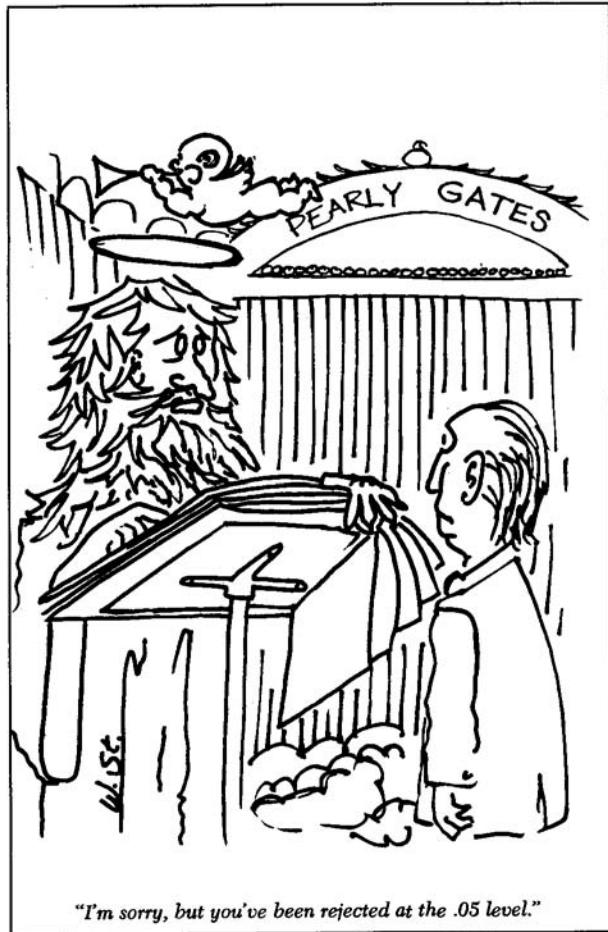
What Is Significant?

An inferential statistical test can tell us whether the results of an experiment would occur frequently or rarely by chance. Inferential statistics with small values occur *frequently by chance*, whereas inferential statistics with large values occur *rarely by chance*. If the result occurs often by chance, we say that it is not significant and conclude that our IV did not affect the DV. In this case we would accept the **null hypothesis**, which says that the differences between groups are due to chance (i.e., not the operation of the IV). If, however, the result of our inferential statistical test occurs rarely by chance (i.e., it is significant), we can conclude that some factor other than chance is operative. If we have conducted our experiment properly and exercised good control (see Chapters 6 and 7), then our significant statistical result gives us reason to believe the IV we manipulated was effective (i.e., did affect the DV scores).

Null hypothesis A hypothesis that says that all differences between groups are due to chance (i.e., not the operation of the IV).

When do we consider that an event occurs rarely by chance? Traditionally, psychologists say that any event that occurs by chance alone 5 times or fewer in 100 occasions is a rare event. Thus, you will see frequent mention of the ".05 level of significance" in journal articles. This statement means that a result is considered significant if it would occur 5 or fewer times by chance in 100 replications of the experiment when the null hypothesis is true. As the experimenter, you decide on the level of significance before the experiment is conducted.

You will encounter several significance tests in later chapters in this book. For the present we will use the *t* test to illustrate their use.



Reprinted by permission of Warren Street.

The *t* Test

For years you have heard the old saying that "clothes make the person." You decide to test this adage experimentally by determining whether type of clothing influences the time it takes a salesperson to wait on customers. To set up the study, imagine that you randomly select 16 salesclerks from a store at the mall and then randomly assign these clerks to one of two groups of eight clerks. Group A will wait on customers in dressy clothes; Group B will wait on customers in sloppy clothes. Because you formed the groups randomly at the start of the experiment, you assume they are comparable before they are exposed to the IV.

The students in Group A wear dressy clothes to the shopping mall, whereas the students in Group B wear sloppy clothes to the shopping mall. Because the students in Group A have no relation to, or effect on, the students in Group B, these groups are *independent* of each other. Each student enters a store in the mall and uses a silent, inconspicuous stopwatch to measure the time (in seconds) it takes a salesperson to offer service. (Keep in mind that these data were *recorded* by the student shoppers; the salesclerks actually produced the

data.) The “latency-to-service” scores (this is a latency DV—see Chapter 6) for the two groups appear below.

Group A (Dressy Clothes)	Group B (Sloppy Clothes)
37	50
38	46
44	62
47	52
49	74
49	69
54	77
69	76
$\Sigma X = 387$	$\Sigma Y = 506$
$M = 48.38$	$M = 63.25$

Do you think the clerks waited on the students in dressy clothes more quickly? Just looking at the differences between the groups suggests that this might be the case; the mean score of Group B is higher than that of Group A. (Higher scores reflect longer times before a salesperson offered service.) On the other hand, there is considerable overlap between the two groups; several of the latency-to-service scores were similar for students dressed in sloppy clothes and the students in the dressy clothes. Is the difference you obtained large enough to be genuine, or is it just a chance happening? Merely looking at the results will not answer that question.

t test An inferential statistical test used to evaluate the difference between two means.

The **t test** is an inferential statistical test used to evaluate the difference between the means of *two groups* (see Chapter 10 for research designs using two groups). Because the two groups in our latency-to-service experiment were independent, we will use an independent-groups *t* test. (We discuss the correlated-groups *t* test in Chapter 10.) The computer printout for our *t* test appears in Table 9-3.

You can see that our *t* value is 2.61 and that the probability of this *t* value is .021. Because the probability of this result occurring by chance is less than .05, we can conclude that the two groups differ significantly.

If your computer program does not provide the probability of your result as part of the printout, you will have to make this determination by yourself. Recall that our *t* value is 2.61.

TABLE 9-3 Computer Printout for Independent-Groups *t* Test

GROUP	<i>N</i>	<i>M</i>	<i>SD</i>
Dressy	8	48.38	9.46
Sloppy	8	63.25	11.73
<i>t</i> = 2.61		<i>df</i> = 14	<i>p</i> = .021

After we have obtained our t value, we must follow several steps in order to interpret its meaning:

- Determine the degrees of freedom (df) involved. (Because some statistical packages may not automatically print the degrees of freedom for you, it is important to keep this formula handy.) For our clothing research:

$$\begin{aligned} df &= (N_A - 1) + (N_B - 1) \\ &= (8 - 1) + (8 - 1) \\ &= 14 \end{aligned}$$

- We use the degrees of freedom (we will discuss the meaning of degrees of freedom after we have completed the problem) to enter a t table (see Table A-1 in Appendix A). This table contains the t values that occur by chance. We will compare our t value to these chance values. To be significant, the calculated t must be equal to or larger than the one given in Table A-1.
- We enter the t table on the row for 14 degrees of freedom. By reading across this row we find that a value of 2.145 occurs by chance 5% of the time (.05 level of significance). Because our value of 2.61 is larger than 2.145 (the .05 value in the table for 14 df), we can conclude that our result is significant (has a probability of occurring by chance less than .05). Thus, the type of clothing had a significant effect on latency to service. This result is one that occurs fewer than 5 times in 100 by chance. Had we chosen a different level of significance, such as once in 100 occurrences (.01), the table value would have been 2.977, and we would have concluded that our result is not significant. In many instances, your computer program will print the probability of your t statistic automatically, and you will not have to consult the t table.

Although it is easy to follow a formula to calculate the degrees of freedom, the meaning of this term may not be clear, even if you have already had an introductory statistics course. We will try to help you understand its meaning. By **degrees of freedom** we mean the ability of a number in a given set to assume any value. This ability is influenced by the restrictions imposed on the set of numbers. For every restriction, one number is determined and will assume a fixed or specified value. For example, assume we have a set of 10 numbers and we know the sum of these numbers to be 100.

Knowing that the sum is 100 is a restriction; hence, one of the numbers will be determined or fixed. In Example 1 below, the last number must be 15 because the total of the first 9 numbers (which can assume any value) is 85. In Example 2, the first 9 numbers have assumed different values. What is the value of the last number?

Degrees of freedom The ability of a number in a specified set to assume any value.

Numbers	1	2	3	4	5	6	7	8	9	10	Sum
Example 1	6	12	11	4	9	9	14	3	17	15	100
Example 2	21	2	9	7	3	18	6	4	5	?	100

As in the first example, the first nine numbers can assume any value. In this example, the sum of the first nine numbers is 75. That means that the value of the last number is fixed at 25.

One-Tail Versus Two-Tail Tests of Significance

Recall from Chapter 5 that you state your experimental hypothesis in either a directional or a nondirectional manner. If you use the directional form, you are specifying exactly how (i.e., the direction) the results will turn out. For the example we have been considering, an experimental hypothesis, stated in general implication form (see Chapter 5), might be as follows:

If students wear dressy clothes to the shopping mall, then the time it takes a salesperson to offer to serve them will be shorter than the latency to service for students dressed in sloppy clothes.

Because we predict that the latency to service for the students wearing dressy clothes will be shorter than that for the students wearing sloppy clothes, we have a directional hypothesis. If we simply indicate that we expect a difference between the two groups and do not specify the exact nature of that difference, then we are using a nondirectional hypothesis.

Now, how do directional and nondirectional hypotheses relate to the t test? If you remember discussing one-tail and two-tail tests of significance in your statistics class, you're on the right track. A one-tail t test evaluates the probability of only one type of outcome, whereas the two-tail t test evaluates the probability of both possible outcomes. If you've associated directional hypotheses with one-tail tests and nondirectional hypotheses with two-tail tests, you're right again.

Figure 9-16 depicts the relation between the type of experimental hypothesis (directional versus nondirectional) and the type of t test used (one-tail versus two-tail). As you can see, the region of rejection is larger and occurs only in one tail of the distribution when a one-tail test is conducted (Figure 9-16A). The probability of the result's occurring by chance alone is split in half and distributed equally to the two tails of the distribution when a two-tail test is conducted (Figure 9-16B).

Although the calculations for a one-tail test of significance and a two-tail test of significance are the same, you would consult different columns in the t table. For the shopping center example, we conducted a two-tail test of significance; 2.145 was our critical value at the .05 level of significance. Hence, a t value equal to or greater than 2.145 is significant (see Table A-1 in Appendix A). Had we done a one-tail test, our critical value at the .05 level of significance would have been 1.761 (see Table A-1).

Because a lower value is required for significance with a one-tail test of significance, it is somewhat easier to find a significant result. If this is the case, why don't experimenters always state directional experimental hypotheses? The main reason is that researchers don't always know exactly how an experiment will turn out. If we knew the outcome of each experiment before it was conducted, there would be no need to do the experiment. If you state a directional hypothesis and then obtain the opposite result, you have to reject your hypothesis. Had you stated a nondirectional hypothesis, your experiment would have confirmed your prediction that there would be a difference between the groups. When conducting a t test, researchers are usually interested in either outcome; for example, what if casually dressed students were actually waited on more quickly?

The Logic of Significance Testing

Remember, we consider the result of an experiment to be statistically significant when it occurs rarely by chance. In such instances we assume that our IV produced the results.

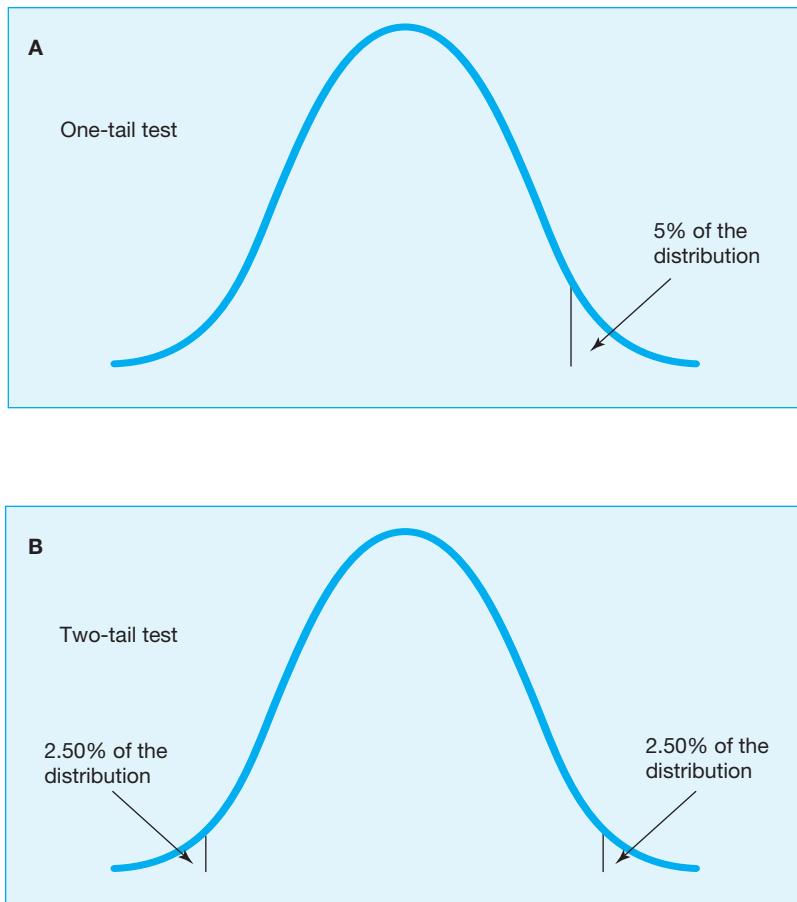


FIGURE 9-16 Regions of Rejection for (A) a One-Tail Test and (B) a Two-Tail Test. **A.** With a one-tail test the region of rejection of the null hypothesis is located in one tail of the distribution. Directional hypotheses, such as $A > B$, are associated with one-tail tests. **B.** With a two-tail test the region of rejection of the null hypothesis is distributed evenly to both tails of the distribution. Nondirectional hypotheses, such as $A \neq B$ (A does not equal B), are associated with two-tail tests.

Although Sherlock Holmes wasn't speaking of a psychological experiment, he captured the intent of significance testing when he asked, "How often have I said to you that when you have eliminated the impossible, whatever remains, *however improbable*, must be the truth?" (Doyle, 1927, p. 111).

Typically our ultimate interest is not in the samples we have tested in an experiment but in what these samples tell us about the population from which they were drawn. In short, we want to generalize, or *infer*, from our samples to the larger population.

We have diagrammed this logic in Figure 9-17. First, samples are randomly drawn from a specified population (Figure 9-17A). We assume that random selection has produced two equivalent groups: Any differences are due solely to chance factors. In Figure 9-17B we see

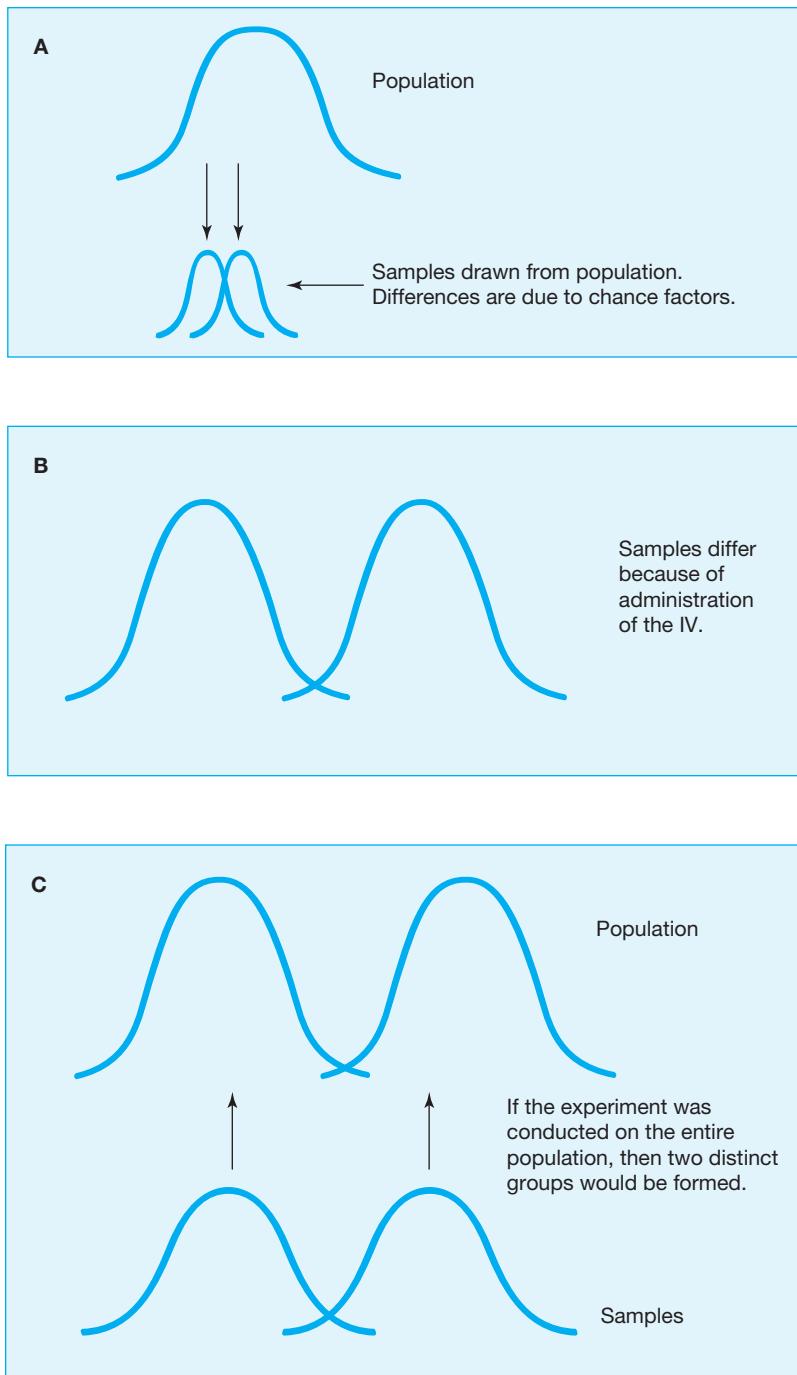


FIGURE 9-17 **A.** Random samples are drawn from a population. **B.** The administration of the IV causes the samples to differ significantly. **C.** The experimenter generalizes the results of the experiment to the general population.

the results of our experiment; the manipulation of the IV caused the groups to be significantly different. At this point generalization begins. Based on the significant difference that exists between the groups, we *infer* what would happen if our treatments were administered to all individuals in the population. In Figure 9-17C we have generalized from the results of our research using two samples to the entire population (see Chapter 8). We are inferring that two separate groups would be created in the population as a result of the administration of our IV.

When Statistics Go Astray: Type I and Type II Errors

Unfortunately, not all our inferences will be correct. Recall that we have determined that an experimental result is significant when it occurs rarely by chance (i.e., 5 times or less in 100). There always is the chance that your experiment represents 1 of those 5 times in 100 when the results did occur *by chance*. Hence, the null hypothesis is true, and you will make an error in accepting your experimental hypothesis. We call this faulty decision a **Type I error** (α). The experimenter directly controls the probability of making a Type I error by setting the significance level. For example, you are less likely to make a Type I error with a significance level of .01 than with a significance level of .05.

On the other hand, the more extreme or critical you make the significance level (e.g., going from .05 to .01) to avoid a Type I error, the more likely you are to make a Type II or beta (β) error. A **Type II error** involves rejecting a *true* experimental hypothesis. Unlike Type I errors, Type II errors are not under the direct control of the experimenter. We can indirectly cut down on Type II errors by implementing techniques that will cause our groups to differ as much as possible. For example, using a strong IV and testing larger groups of participants are two techniques that will help avoid Type II errors.

We will have more to say about Type I and Type II errors in subsequent chapters. They are summarized here as follows:

		TRUE STATE OF AFFAIRS	
		Experimental Hypothesis Is True	Null Hypothesis Is True
YOUR DECISION	Experimental Hypothesis Is True	Correct Decision	Type I (α) Error
	Null Hypothesis Is True	Type II (β) Error	Correct Decision

You should keep in mind that the typical practice is to set the alpha level at .05 because it places the probability of Type I and II errors at acceptable levels.

Type I error Accepting the experimental hypothesis when the null hypothesis is true.

Type II error Accepting the null hypothesis when the experimental hypothesis is true.

Effect Size

Effect size The magnitude or size of the experimental treatment.

Before concluding this chapter, we want to introduce you to a statistical concept and procedure that currently is gaining in popularity and frequency of usage. **Effect size** is a statistical measure that conveys information concerning the *magnitude* of the effect produced by the IV.



Doesn't obtaining significance with an inferential test give us the same information? After all, significance indicates the IV had an effect, and that's our concern. Why do we need anything else?

Unfortunately, a significant statistical test tells us only that the IV had an effect; it does not tell us about the size of the significant effect. Moreover, whether an effect is significant or not may depend on factors other than the IV. For example, you just saw that you are more likely to obtain significance (i.e., avoid a Type II error) when you use larger samples, even though the influence of the IV remains the same. The American Psychological Association's *Publication Manual* (2001) indicates that "neither of the two types of probability values [your selected alpha level and the probability level associated with the inferential statistic you calculate] reflects the importance or magnitude of an effect because both depend on sample size" (p. 18).

Such considerations have encouraged researchers to report effect size in addition to the inferential statistics that are used. In fact, some statisticians (e.g., Kirk, 1996) envision a time when the reporting of effect size will be more common than significance testing. Indeed, the APA's *Publication Manual* (2001) states, "You are encouraged to provide effect-size information" (p. 18) when you prepare your research report.

There are several different ways to calculate effect size. Here are two that should give you no problems. Cohen's d (Cohen, 1977) is easy to compute when you use two groups and calculate a t test. In these cases:

$$d = \frac{t(N_1 + N_2)}{\sqrt{df} \sqrt{N_1 N_2}}$$

or, when the two samples are of equal size:

$$d = \frac{2t}{\sqrt{df}}$$

Cohen (1977) indicated that $d = .20$ to $.50$ is a small effect size, $d = .50$ to $.80$ is a medium effect size, and d values greater than $.80$ reflect large effect sizes.

A second technique for determining effect size is appropriate when you calculate a Pearson product-moment correlation (r): r^2 gives you an estimate of the proportion of the variance accounted for by the correlation in question (Rosenthal & Rosnow, 1984). For example, even though $r = .30$ is significant ($p < .01$) with 90 pairs of scores, this correlation accounts for only 9% ($.30^2 = .09 = 9\%$) of the variance. This figure means that 91% of the variability in your research results is accounted for by *other* variables, a rather small effect size indeed.

■ REVIEW SUMMARY

1. A **correlation coefficient** is a single number that represents the degree of relationship between two variables. Many predictions are based on correlations.
2. A perfect negative correlation (-1) exists when an increase of 1 unit in one variable is always accompanied by a proportional decrease in the other variable. A perfect positive correlation ($+1$) exists when an increase of 1 unit is always accompanied by a proportional increase in the other variable. A correlation of 0 indicates that there is no relation between the variables under consideration.
3. The Pearson product-moment correlation coefficient is calculated when both variables are interval-scale measurements.
4. Inferential statistics help the experimenter decide whether the IV was effective. A significant inferential statistic is one that occurs rarely by chance.
5. The **t test**, which is an inferential statistic, is used to test the differences between two groups.
6. When results are significant, the experimenter hopes to be able to extend the results of the experiment to the more general population.
7. A one-tail *t* test is conducted when a directional hypothesis is stated, whereas a two-tail *t* test is conducted when a nondirectional hypothesis is stated.
8. Even though lower critical values are associated with one-tail tests, making it easier to attain significance, most experimental hypotheses are nondirectional because the researchers do not know exactly what the research will show.
9. Sometimes the results of an inferential statistical test produce an incorrect decision. An experimental hypothesis may be incorrectly accepted (**Type I error**) or incorrectly rejected (**Type II error**).

■ Check Your Progress

1. Matching
 1. correlation coefficient
 2. perfect negative correlation
 3. perfect positive correlation
 4. significant
 5. inferential statistics
 6. Type I error
 7. Type II error
 8. one-tail test
 9. two-tail test
 - A. nondirectional hypothesis
 - B. result occurs infrequently by chance
 - C. rejecting a true null hypothesis
 - D. tests conducted to determine whether the IV had an effect
 - E. directional hypothesis
 - F. represents the degree of relationship between two variables
 - G. rejecting a true experimental hypothesis
 - H. -1
 - I. $+1$
2. Explain the difference between a positive correlation and a perfect positive correlation.

3. What does a zero correlation signify?
4. Explain the logic involved when an independent-groups t test is conducted.
5. What is meant by “level of significance”? How is the level of significance determined?
6. If it is easier to obtain a significant result with a one-tail test, why would an experimenter ever state a nondirectional experimental hypothesis and thus use a two-tail test?
7. A one-tail test of significance is associated with a
 - a. directional hypothesis
 - b. nondirectional hypothesis
 - c. positive correlation
 - d. negative correlation
8. The Type I error
 - a. is under the direct control of the experimenter
 - b. always occurs 5% of the time
 - c. is specified by the experimental hypothesis
 - d. all of the above
 - e. none of the above
9. If you could compare all men and women in the world, you would find that men are significantly more aggressive. You conduct an experiment and find no difference in aggression between men and women. You have made a
 - a. correct decision
 - b. Type I error
 - c. Type II error
 - d. Type III error

■ Key Terms

Statistics, 171	Pie chart, 176	Normal distribution, 184
Descriptive statistics, 171	Histogram, 177	Correlation coefficient, 188
Inferential statistics, 171	Bar graph, 177	Null hypothesis, 192
Measurement, 172	Frequency polygon, 178	t test, 194
Scale of measurement, 172	Line graph, 178	Degrees of freedom, 195
Nominal scale, 172	Ordinate, 178	Type I error, 199
Ordinal scale, 172	Abscissa, 178	Type II error, 199
Interval scale, 172	Variability, 182	Effect size, 200
Ratio scale, 173	Range, 182	
Mode, 174	Variance, 182	
Median, 174	Standard deviation, 183	
Mean, 174		

■ Looking Ahead

So far we have considered sources of researchable problems (Chapter 1), developed an experimental hypothesis (Chapter 2), considered the ethics involved in conducting research (Chapter 2), scrutinized our experiment for possible extraneous variables and nuisance variables (Chapter 6), and implemented control procedures to deal with these extraneous variables (Chapters 6 and 7). Now we are ready to combine all of these elements in an experimental design. In Chapter 10 experimental designs involving the use of two groups of participants are considered. We will consider more complex designs in subsequent chapters.

Designing, Conducting, Analyzing, and Interpreting Experiments with Two Groups

Experimental Design: The Basic Building Blocks

- The Two-Group Design
- Comparing Two-Group Designs
- Variations on the Two-Group Design

Statistical Analysis: What Do Your Data Show?

- The Relation Between Experimental Design and Statistics

- Analyzing Two-Group Designs
- Calculating Your Statistics

Interpretation: Making Sense of Your Statistics

- Interpreting Computer Statistical Output

The Continuing Research Problem

Experimental Design: The Basic Building Blocks

Now that the preliminaries are out of the way, we are ready to begin an experiment. Or are we? Although we have chosen a problem, read the relevant literature, developed a hypothesis, selected our variables, instituted control procedures, and considered the participants, we are still not ready to start the experiment. Before we can actually begin, we must select a blueprint. If you were about to design a house, you would be faced with an overwhelming variety of potential plans—you would have many choices and selections ahead of you. Fortunately, selecting a blueprint for your experiment is simpler than designing a house because there are relatively few standard choices used by experimenters in designing their experiments.

Selecting a blueprint for your experiment is just as important as selecting one for a house. Can you imagine what a house would look like if you began building it without any plans? The result would be a disaster. The same is true of “building” an experiment. We refer to the research blueprint as our **experimental design**.

In Chapter 1 you learned that an experimental design is the general plan for selecting participants, assigning those participants to experimental conditions, controlling extraneous variables, and gathering data. If you begin your experiment without a proper design, your experiment may “collapse” just as a house built without a blueprint might. How can an experiment collapse? We have seen students begin experiments without any direction only to end up with data that fit no known procedure for statistical analysis. We also have seen students collect data that have no

Experimental design

The general plan for selecting participants, assigning participants to experimental conditions, controlling extraneous variables, and gathering data.

bearing on their original question. Thus, we hope not only that you will use this text during your current course but also that you will keep the book and consult it as you design research projects in the future.

In this chapter we will begin developing a series of questions in a flowchart to help you select the correct design for your experiment. As Charles Brewer, distinguished professor of psychology at Furman University, is fond of saying, "If you do not know where you are going, the likelihood that you will get there borders on randomness" (Brewer, 2002, p. 503). If you don't design your experiment properly, the probability that it will answer your research question is slim. Sherlock Holmes knew this lesson well: "No, no: I never guess. It is a shocking habit—destructive to the logical faculty" (Doyle, 1927, p. 93).

When you were a child and played with Legos or Tinkertoys, you probably got a beginner's set first. This set was small and simple, but with it you learned the basics of building. As you got older, you could use larger sets that allowed you to build and create more complicated objects. The parallel between children's building sets and experimental design is striking. In both cases the beginner's set helps us learn about the processes involved so that we can use the advanced set later; the basic set forms the backbone of the more advanced set. In both cases, combining simple models increases the possibilities for building, although more complex models must still conform to the basic rules of building.

The Two-Group Design

In this chapter we examine the most basic experimental design—the two-group design—and its variations. This design is the simplest possible one that can yield a valid experiment. In research situations, we typically follow the **principle of parsimony**, also known as Occam's (or Ockham's) razor. William of Occam, a fourteenth-century philosopher, became famous for his dictum "Let us never introduce more than is required for an explanation" (McInerny, 1970, p. 370). In research, we apply the principle of parsimony to research questions, just as detectives apply the principle of parsimony to their investigations: Don't needlessly complicate the question that you are asking. The two-group design is the most parsimonious design available.

Principle of parsimony

The belief that explanations of phenomena and events should remain simple until the simple explanations are no longer valid.

Independent variable (IV)

A stimulus or aspect of the environment that the experimenter directly manipulates to determine its influences on behavior.

Dependent variable (DV)

A response or behavior that the experimenter measures. Changes in the DV should be caused by manipulation of the independent variable (IV).

How Many IVs? Figure 10-1 shows the first question we must ask in order to select the appropriate design for our experiment: "How many **independent variables (IVs)** will our experiment have?" In this chapter and the next we will deal with experimental designs that have one IV. You will remember (see Chapter 6) that an IV is a stimulus or aspect of the environment that the experimenter directly manipulates to determine its influences on behavior, which is the **dependent variable (DV)**. If you want to determine how anxiety affects test performance, for example, anxiety would be your IV. If you wish to study the effects of different therapies on depression, the different therapies would be your IV. The simplest experimental design has only one IV. We will look at research designs with more than one IV in Chapter 12.

A minority of published research studies use one IV. Does that mean that experiments with one IV are somehow poor or deficient? No, there is nothing wrong with a one-IV design; however, such a design is simple and may not yield

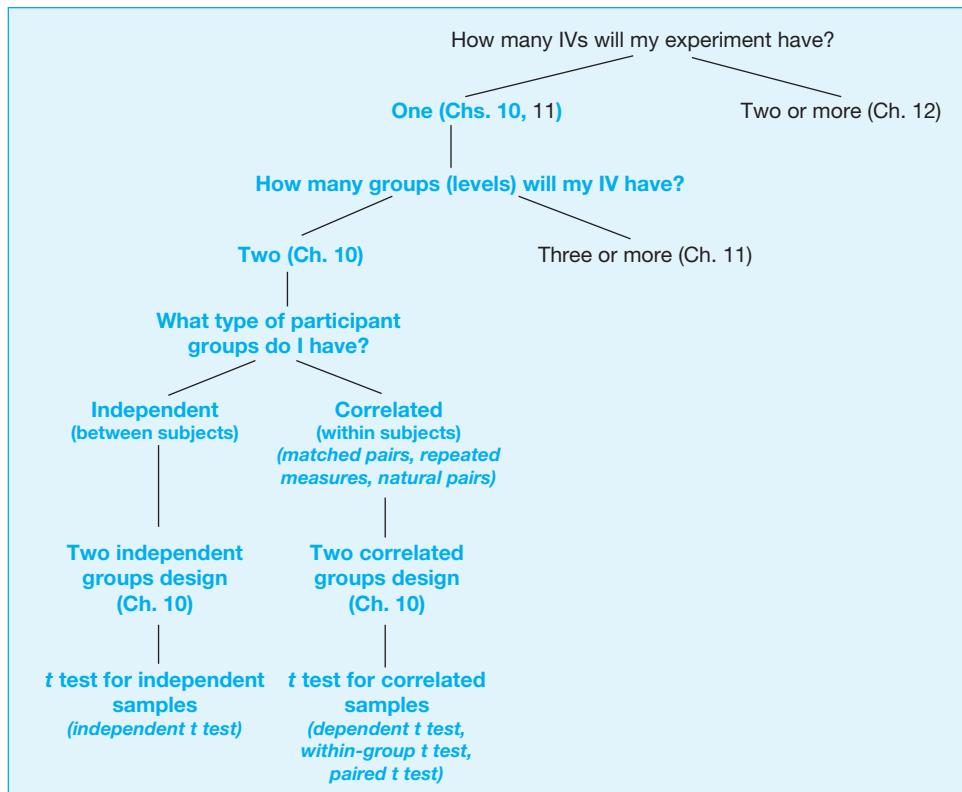


FIGURE 10-1 Experimental Design Questions.

all the answers an experimenter desires. Simple is not necessarily bad. Inexperienced researchers or researchers who are beginning to investigate new areas often prefer single-IV designs because they are easier to conduct than multiple-IV designs, and it is simpler to institute the proper control procedures. Also, the results of several one-IV experiments, when combined in one report, can describe complex phenomena.

How Many Groups? Assuming we have chosen to use a single-IV design, we come to our second question (see Figure 10-1) in determining the proper experimental design: “How many groups will I use to test my IV?” In this chapter the answer is two. Although an experiment can have a single IV, it must have at least two groups.



Why must we have two groups but only one IV?

The simplest way to determine whether our IV caused a change in behavior is to compare some research participants who have received our IV to some others who have not received the IV. If those two groups differ, and we are assured that we controlled potential **extraneous variables** (see Chapter 6), then we can conclude that the IV caused the participants to differ. The way we can test two groups with only one IV is to make the two groups differ in the amount or the type of the IV that they receive. Note carefully that the last statement is *not* the same as saying that the groups have different IVs.

Extraneous variables

Uncontrolled variables that may unintentionally influence the dependent variable (DV) and thus invalidate an experiment.

Levels

Differing amounts or types of an IV used in an experiment (also known as *treatment conditions*).

The most common manner of creating two groups with one IV is to present some amount or type of IV to one group and to withhold that IV from the second group. Thus, the experimenter contrasts the *presence* of the IV with the *absence* of the IV. These differing amounts of the IV are referred to as the **levels** (also known as *treatment conditions*) of the IV. Thus, in the common two-group design, one level of the IV is none (its absence) and the other is some amount (its presence). Notice that the presence and absence of an IV is conceptualized as two differing levels of the same IV rather than as two different IVs. Now let's return to our earlier examples.



If you were interested in the effects of anxiety on test performance or the effects of therapy on depression, how would you implement an experimental design so that you could compare the presence of the IV to the absence of the IV?

In the first example you would have to compare anxious test takers (first level) to nonanxious test takers (second level). In the second example you would compare depressed people receiving therapy (first level) to depressed people who were not receiving therapy (second level).

Experimental group

In a two-group design, the group of participants that receives the IV.

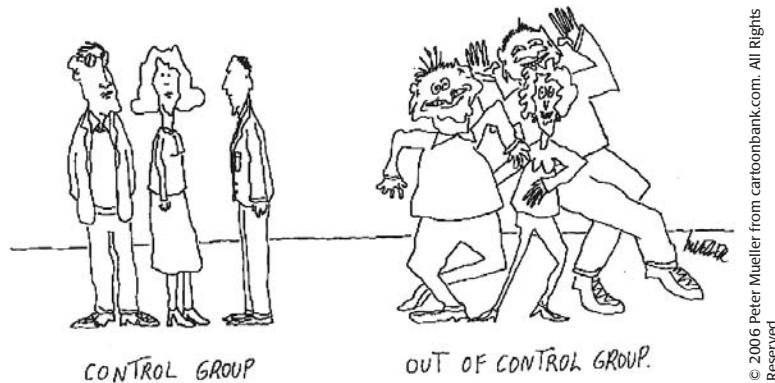
Control group

In a two-group design, the group of participants that does not receive the IV.

In this presence-absence situation, the group of participants receiving the IV is typically referred to as the **experimental group**. It is as if we were conducting the experiment on them. The group that does not receive the IV is known as the **control group**. The members of this group serve a control function because they give us an indication of how people or animals behave under "normal" conditions—that is, without being exposed to our IV. They also serve as a comparison group for the experimental group. We will use statistics to compare performance scores on the DV for the two groups to determine whether the IV has had an effect. When our two groups differ significantly, we assume that the difference is due to the IV. If the groups do not differ significantly, we conclude that our IV had no effect.

Let's look at a research example using the two-group design. Kristen McClellan, a student at Madonna University in Livonia, Michigan, and her faculty advisor,

Edie Woods, were interested in how salesclerks reacted to customers with a disability. Their IV was the disability of hearing loss. Their experimental group of salesclerks thus encountered pairs of customers who were deaf (easily identifiable because they entered the store conversing in sign language), and the control group of clerks waited on hearing pairs of customers. McClellan and Woods (2001) randomly assigned 77 salesclerks to either the experimental or



© 2006 Peter Mueller from cartoonbank.com. All Rights Reserved.

Fortunately, we deal with control groups and experimental groups in psychological research!

control group and had the confederates (customers) enter the store. The confederates unobtrusively timed how long it took a salesperson to approach and offer assistance after initial eye contact. (They did not obtain informed consent from the clerks because the clerks were participants at minimal risk; see Chapter 2.) McClellan and Woods found that the clerks in the experimental group took significantly longer on average (3.9 minutes) to wait on the deaf customers than the clerks in the control group, who took 1.3 minutes to wait on the hearing customers. This basic two-group design is depicted in Figure 10-2. Researchers frequently use block diagrams, such as the one shown in Figure 10-2, to portray the design of an experiment graphically. Note that the IV (customer hearing) heads the entire block; the two levels of the IV comprise the two subdivisions of the block. We will use this building-block notation throughout the three chapters concerning experimental design so that you can conceptualize the various designs more easily.

Assigning Participants to Groups We have yet another question to face before we can select our experimental design: We must decide how we plan to assign our research participants to groups. (Before the 1994 edition, the American Psychological Association's *Publication Manual* referred to research participants as "subjects." You are likely to read this term in older research studies; you may also hear the term for some time before the psychological language becomes standardized—it's hard to teach old psychologists new tricks!) We can assign our participants either randomly or in some nonrandom fashion. We will examine random assignment first.

Random Assignment to Groups As you saw in Chapter 6, an often-used approach for assigning participants to groups is **random assignment**. Given that we are dealing with only two groups, we could flip a coin and assign participants to one group or the other on the basis

Random assignment

A method of assigning research participants to groups so that each participant has an equal chance of being in any group.

INDEPENDENT VARIABLE (CUSTOMER HEARING)

EXPERIMENTAL GROUP	CONTROL GROUP
--------------------	---------------

Customers were deaf	Customers were not deaf
---------------------	-------------------------

FIGURE 10-2 The Basic Two-Group Design.

Random selection A control technique that ensures that each member of the population has an equal chance of being chosen for an experiment.

Independent groups
Groups of participants formed by random assignment.

Between-subjects comparison Refers to a contrast between groups of participants who were randomly assigned to groups.

of heads or tails. As long as we flip a fair coin, our participants would have a 50-50 chance of ending up in either group. Remember that random assignment is *not* the same as **random selection**, which we also learned about in Chapter 6. Random selection deals with choosing your research participants, whereas random assignment refers to putting those participants into their groups. Random assignment is concerned with control procedures in the experiment, whereas random selection influences the generality of the results. We examined the generality issue more closely in Chapter 8.

When we randomly assign our participants to groups, we create what are known as **independent groups**. The participants in one group have *absolutely no* ties or links to the participants in the other group; they are independent of each other. If you tried to relate or pair a participant in one group to one in the other group, there would be no logical way to do so. When we wish to compare the performance of participants in these two groups, we are making what is known as a **between-subjects comparison**. We are interested in the difference *between* these two groups of participants who have no ties or links to each other.

Terminology in experimental design can sometimes be confusing—for some reason, it has never been standardized. You may hear different people refer to *independent groups designs*, *randomized groups designs*, or *between-subjects designs*. All these names refer to the same basic strategy of randomly assigning participants to groups. The key to avoiding confusion is to understand the principle behind the strategy of random assignment. When you understand the basic ideas, the names will make sense. As you saw in Figure 10-1, when we have one IV with two levels and participants who are assigned to groups randomly, our experiment fits the two-independent-groups design.

Random assignment is important in experimental design as a control factor. When we randomly assign our research participants to groups, we can usually assume that our two groups will now be equal on a variety of variables (Spatz, 2001). Many of these variables could be extraneous variables that might confound our experiment if left uncontrolled (see Chapter 6). Random assignment is one of those statistical procedures that is *supposed* to work—in the long run. There are no guarantees that it will create the expected outcome, however, if we select a small sample. For example, we would not be surprised to flip a coin 10 times and get 7 or 8 heads. On the other hand, if we flipped a coin 100 times, we would be quite surprised if we obtained 70 or 80 heads.

Thus, when we randomly assign participants to groups, we expect the two groups to be equal on a variety of variables that could affect the experiment's outcome. Think back to McClellan and Woods's (2001) experiment dealing with customers with and without hearing disability and salesclerks' response times. When we described their experiment, we were careful to point out that the researchers *randomly* assigned some clerks to wait on deaf customers and *randomly* assigned some clerks to wait on hearing customers. What if they had picked the most polite clerks to wait on the hearing customers (control group)? Putting the polite clerks in the control group would cause the clerks' politeness to vary systematically with levels of the IV (that is, polite clerks would wait on hearing customers, but less-polite clerks would not). Such assignment to groups would result in a **confounded experiment** (see

Confounded experiment
An experiment in which an extraneous variable varies systematically with the IV, which makes drawing a cause-and-effect relation impossible.

Chapter 6). If the results showed that the control-group clerks waited on the hearing customers faster than the experimental-group clerks waited on deaf customers, we could not draw a definite conclusion about why that result happened. They might have responded more quickly because they were more polite, because they were waiting on hearing customers, or because of the combination of these two factors. Unfortunately, with a confounded experiment there is no way to determine which conclusion is appropriate. If McClellan and Woods had conducted their experiment in this manner, they would have wasted their time.

Let us remind you of one more benefit of random assignment. In Chapter 6 you learned that random assignment is the only technique we have that will help us control unknown extraneous variables. For example, in McClellan and Woods's (2001) experiment, what extraneous variables might affect salesclerks' performance? We have already identified the clerks' politeness as a possibility. The researchers did not have access to any politeness measure for the clerks. Other variables not even considered could also play a role in the clerks' performance; therefore, McClellan and Woods were careful to assign their clerks randomly to experimental and control groups. Random assignment should equate any differences between the two groups.



Can you think of a flaw in McClellan and Woods's reasoning behind random assignment?

Random assignment is a technique that *should* work in the long run. Because McClellan and Woods assigned 77 clerks to the two groups, there is a good chance that random assignment made the groups perfectly equal. If they had measured only a few salesclerks, however, random assignment might not have created equal groups. If you thought of this potential problem, congratulations! What can we do when we conduct an experiment with small numbers of participants?

Nonrandom Assignment to Groups In the previous section we saw a potential pitfall of random assignment: The groups may not be equal after all. If we begin our experiment with unequal groups, we have a problem. Remember that random assignment should create equal groups in the long run. In other words, as our groups get larger, we can place more confidence in random assignment achieving what we want it to.

Suppose we are faced with a situation in which we have few potential research participants and we are worried that random assignment may not create equal groups. What can we do? In this type of situation, we can use a nonrandom method of assigning participants to groups. What we will do is either capitalize on an existing relationship between participants or create a relationship between them. In this manner, we know something important about our participants before the experiment, and we will use **correlated assignment** (also known as *matched* or *paired assignment*) to create equal groups. Thus, we use correlated assignment whenever there is a relationship between the participants in the groups.

Correlated assignment

A method of assigning research participants to groups so that there is a relationship between small numbers of participants; these small groups are then randomly assigned to treatment conditions (also known as *paired* or *matched assignment*).

(Be careful—correlated assignment has *nothing* to do with computing a correlation coefficient.) There are three common ways to use correlated assignment.

Matched pairs

Research participants in a two-group design who are measured and equated on some variable before the experiment.

1. **Matched Pairs.** To create **matched pairs**, we must measure our participants on a variable (other than our IV) that could affect performance on our experiment's DV. Typically we measure a variable that could result in confounding if not controlled. After we have measured this variable, we create pairs of participants that are equal on this variable. After we have created our matched pairs, we then randomly assign participants from these pairs to the different treatment conditions.

If this description seems confusing, an example should help clarify matters. Imagine that we wanted to replicate McClellan and Woods's salesclerk study because we were worried that random assignment may not have created equal groups. Suppose we suspect that female salesclerks tend to wait on customers faster than male clerks (probably a totally fictitious supposition, but it makes a good example). In this situation we would be concerned if the ratio of female to male clerks differed between the groups. If we flip a coin to assign clerks to groups, the sex ratio of the two groups might be unequal; therefore, we decide to use matched assignment to groups. First, we create our matched pairs. The first pair consists of two clerks of the same sex; the second pair is another two clerks of the same sex. (We pair all the other clerks by sex also.) For each pair, we flip a coin to determine which clerk to assign to the experimental group and which clerk to assign to the control group. Then we repeat the procedure for our second pair, and so on. After we complete this matched assignment, we have an experimental group and a control group that are perfectly balanced in terms of sex. We have used matched assignment to create equal groups before our experiment begins. (Note: In this hypothetical example, if there were an odd number of clerks, one clerk would not have a match, and we could not use that clerk in the experiment.)



In what way is matched assignment guaranteed to create equal groups when random assignment is not?

The beauty of matched assignment is that we have measured our participants on a specific variable that could affect their performance in our experiment, and we have equated them on that variable. When we use random assignment, we are leaving this equating process to chance. Remember, we must match on a variable that could affect the outcome of our experiment. In the fictitious example we just used, you would have to be certain that sex was linked to salesclerks' performance. If you cannot measure your participants on a variable that is relevant to their performance in your experiment, then you should not use matched assignment. If you match your participants on a variable that is not relevant to their performance, then you have actually hurt your chances of finding a significant difference in your experiment (see "Statistical Issues" in the "Advantages of Correlated Group Designs" section later in this chapter).

2. **Repeated Measures.** In **repeated measures**, we use the ultimate in matched pairs: We simply test or measure the same participants in both treatment conditions of our experiment. The matched pairs here are perfectly equal because they consist of the same people or animals tested across the entire experiment. No extraneous variables should be able to confound this situation because any difference between the participants' performance in the two treatment conditions is due to the IV. In this type of experiment, participants serve as their own controls.

Repeated measures An experimental procedure in which research participants are tested or measured more than once.



Why is it not possible to use repeated measures in all experiments? Try to think of two reasons.

Thinking about using repeated measures for our groups forces us to consider some practical factors:

- a. Can we remove the effects of the IV? McClellan and Woods (2001) did not do anything to the salesclerks that had lasting effects—waiting on deaf customers should not affect how clerks respond to other customers in the future. In a study cited in Chapter 2, Burkley et al. (2000) could not use repeated measures in their experiment because they could not remove the effects of reading a particular consent form on the students in their experiment. Think about it carefully—even though they could have allowed students to read the second form, they could not have removed the effects of reading the other form first. If the students remembered the previous form, giving them a new form would not remove its effects. Sometimes when researchers cannot remove the effects of an IV, they deal with this problem by using counterbalancing (see Chapter 6). By balancing the effects across groups, researchers hope that the effects end up equal across those groups.
- b. Can we measure our DV more than once? To this point we have not focused on the DV in this chapter because it has little to do with the choice of an experimental design. When you consider using repeated measures, however, the DV is extremely important. When you use repeated measures, the DV is measured multiple times (at least twice). McClellan and Woods could have used repeated measures by having each clerk wait on both a deaf and a hearing customer. In some cases, however, it is simply not possible to use the same DV more than once. In Burkley et al.'s (2000) experiment, as soon as the students had solved the anagrams, solving the anagrams could not be used again as a DV. In other cases, we may be able to use a similar DV, but we must be cautious. To use repeated measures on solving anagrams in Burkley et al.'s experiment, we would need two different sets of anagrams, one for testing students with each type of consent form. If we use two different forms of our DV, we must ensure that they are comparable. Although we could use two sets of anagrams, they would have to be equally difficult, which

might be hard to determine. The same would be true of many different DVs such as mazes, tests, puzzles, and computer programs. If we cannot assure comparability of two measures of the same DV, then we should not use repeated-measures designs.

- c. Can our participants cope with repeated testing? This question relates at least to some degree to the ethics of research, which we covered in Chapter 2. Although no specific ethical principles speak to the use of repeated measures, we should realize that requiring extended participation in our experiment might affect participants' willingness to take part in the research. Also, in extreme cases extended time in a strenuous or taxing experiment could raise concerns for physical or emotional well-being. Another of our worries in this area is whether human participants will agree to devote the amount of time that we request of them in a repeated-measures design.

It is important that we think about these practical considerations when weighing the possibility of a repeated-measures design. Although repeated-measures designs are one of our better control techniques, there are some experimental questions that simply do not allow the use of such a design.

Natural pairs Research participants in a two-group design who are naturally related in some way (e.g., a biological or social relationship).

3. **Natural pairs.** **Natural pairs** are essentially a combination of matched pairs and repeated measures. In this technique we create pairs of participants from naturally occurring pairs (e.g., biologically or socially related). For example, psychologists who study intelligence often use twins (natural pairs) as their research participants. This approach is similar to using the same participant more than once (repeated measures), but it allows you to compose your pairs more easily than through matching. Thus, when an experiment uses siblings, parents and children, husbands and wives, littermates, or some other biological or social relationship, that experiment has used natural pairs.

In summary, whenever there is a relationship between participants in different groups, a correlated-groups design is being used. By looking at Figure 10.1, you can see that an experiment with (a) one IV that has (b) two levels, in which you plan to use (c) correlated assignment of participants to groups, results in the *two-correlated-groups design*. Participants who have been matched on some variable or who share some relationship would have scores that

are related. When we wish to compare the performance of such participants, we are making what has traditionally been known as a **within-subjects comparison**. We are essentially comparing scores within the same participants (subjects). Although this direct comparison is literally true only for repeated-measures designs, participants in matched or natural pairs are the same with regard to the matching variable.

Let's look at a student example of a two-correlated-groups design. A major area of study in psychology is the effect of stress on the body and the body's reactions. Rachel Wells (2001), a student from Nebraska Wesleyan University,

found that previous researchers had used a mental arithmetic test to induce stress in participants. She wanted to determine the effects of such a test on college students' bodily reactions. She had students count backward from 715 by 13, telling them that most students could complete the task in 4 minutes. Immediately after counting for 4 minutes, she measured the participants' heart rate and blood pressure. Students then spent 10 minutes

Within-subjects comparison Refers to a contrast between groups of participants who were assigned to groups through matched pairs, natural pairs, or repeated measures.

completing questionnaires as a nonstressful rest period. After the 10-minute period, Wells measured the participants' heart rate and blood pressure again. The students showed a decrease in both heart rate and blood pressure, demonstrating that the mental arithmetic was indeed stress provoking.

Because she used repeated measures, Wells's experiment is a good example of a correlated-groups design. She measured the students' body signs after a stressful event and then again after a rest period. The measurement after the stressor was a posttest; the measurement after rest served as the comparison period. Often in such research, the experimenter might measure the body reactions *before* inducing the stressor; this measurement would be a pretest. In Wells's experiment, the IV was the stress induced, and the DVs were the students' physiological reactions to the stress. Some of the particularly important participant (subject) variables, which Wells controlled by the use of repeated measures, were participants' ages (perhaps younger or older people have different reactions to stress), time of day (perhaps physiological indicators vary by the time of day), and students' intelligence (perhaps brighter students would find the task less stressful). All these extraneous variables (and others) were controlled because the *same* students took both the comparison test and the posttest.

What if Wells had wanted to use matched pairs rather than repeated measures? Remember that matching should occur on a relevant variable—one that could be an extraneous variable if left unchecked. Suppose that these students varied widely in their math ability. Wells might have wanted to create matched pairs based on that ability, thinking that math ability would likely affect performance on the counting task, which could determine the level of stress felt by each student. Wells could have used other variables for matching as long as she was certain that the variables were related to the students' physiological indicators.

Could Wells have run this experiment using natural pairs? Based on the information given you, there is no indication that natural pairs of students existed. If the students were sets of twins, then the experiment would be ideally suited for natural pairs. It seems unlikely that there was any factor that made these students natural pairs, so if pairing was important to Wells, she would have had to create pairs by matching on some variable.

■ REVIEW SUMMARY

1. Psychologists plan their experiments beforehand using an **experimental design**, which serves as a blueprint for the experiment.
2. The two-group design applies to experimental situations in which one IV has two levels or conditions.
3. Two-group designs often use an **experimental group**, which receives the IV, and a **control group**, which does not receive the IV.
4. **Randomly assigning** research participants to groups results in **independent groups** of participants.
5. **Correlated groups** of research participants are formed by creating **matched pairs**, using **natural pairs**, or by measuring the participants more than once (**repeated measures**).

Check Your Progress

1. Why can't we conduct a valid experiment with only one group?
2. The differing amounts of your IV are known as the _____ of the IV.
3. How are independent groups and correlated groups different? Why is this difference important to experimental design questions?
4. Matching

1. random assignment	A. brother and sister
2. natural pairs	B. take the same test each month
3. repeated measures	C. two people with the same IQ
4. matched pairs	D. flipping a coin
5. In what type of situation do we have to be most careful when using random assignment as a control technique?
6. You are planning an experiment that could use either independent groups or correlated groups. Under what conditions should you use a correlated-groups design? When is it acceptable to use random assignment?
7. Random assignment is more likely to create equal groups when

a. small samples are involved	c. a directional hypothesis is being tested
b. large samples are involved	d. a nondirectional hypothesis is being tested

Comparing Two-Group Designs

Because there are two different two-group designs, researchers must choose whether they want to design their experiment with independent or correlated groups. You may be wondering how researchers make such a choice. In the next sections we will cover some issues that psychologists must consider when they plan their research studies. Read carefully—you may be facing this choice yourself in the future.

Look at Figure 10-1 again. You can see that the two-independent-groups design and the two-correlated-groups design are quite similar. Both designs describe experimental situations in which you use one IV with two groups. The only difference comes from how you assign your participants to groups. If you simply assign on a random basis, you use the two-independent-groups design. On the other hand, if you match your participants on some variable, if you test your participants twice, or if your participants share some relationship, you use the two-correlated-groups design.

Choosing a Two-Group Design Now that you have two experimental designs that can handle very similar experimental situations, how do you choose between them? Should you use independent groups, or should you use correlated groups of some sort?

You may remember we said that random assignment is supposed to “work” (i.e., create equal groups) in the long run. If you are using large groups of participants, therefore, random assignment should equate your groups adequately. The next question, of course, is how large is large? Unfortunately, there is no specific answer to this question—the answer may vary from researcher to researcher. If you are using 20 or more participants per group, you can

feel fairly safe that randomization will create equal groups. On the other hand, if you are using 5 or fewer participants in a group, randomization may not work. Part of the answer to the question of numbers boils down to what you feel comfortable with, what your research director feels comfortable with, or what you think you could defend to someone. In McClellan and Woods's (2001) study, there were 77 salesclerks divided into two groups. Given our guidelines, this number is quite adequate to use random assignment to create independent groups. On the other hand, although Wells (2001) had 41 students participate in the experiment, she may have decided to use repeated measures because she was worried about individual levels of stress that could have created a great deal of variability. Whatever you decide, it is critical to remember that the larger your samples, the more likely random assignment is to create equal groups.

Advantages of Correlated-Groups Designs There are two primary advantages correlated-groups designs provide to researchers: control and statistical issues. Both advantages are important to you as an experimenter.

Control Issues One basic assumption that we make before beginning our experiment is that the participants in our groups are equal with respect to the DV. When our groups are equal before the experiment begins and we observe differences between our groups on the DV after the experiment, then we can attribute those differences to the IV. Although randomization *should* equate our groups, the three methods for creating correlated-groups designs give us greater certainty of equality. We have exerted control to create equal groups. Thus, in correlated designs, we have some "proof" that our participants are equal beforehand. This equality helps us reduce some of the error variation in our experiment, which brings us to the statistical issues.

Statistical Issues Correlated-groups designs can actually benefit us statistically because they can help reduce error variation. You might be wondering, "What is error variation, anyhow?" In an experiment that involves one IV, you essentially have two sources of variability in your data. One source of variation is your IV: Scores on the DV should vary due to the two different treatment groups you have in your experiment. This source of variation, referred to as **between-groups variability**, is what you are attempting to measure in the experiment. Other factors that can cause variation in the DV, such as individual differences, measurement errors, and extraneous variation, are collectively known as **error variability**. As you might guess, our goal in an experiment is to *maximize* the between-groups variability and *minimize* the error or within-groups variability.

Why is it important to reduce error variability? Although formulas for different statistical tests vary widely, they all reduce to the following general formula:

$$\text{statistic} = \frac{\text{between-groups variability}}{\text{error variability}}$$

Remember that the probability of a result occurring by chance goes down as the value of your statistic increases. Thus, larger statistical values are more likely to show significant differences in your experiment. Your knowledge of math tells you that there are two ways to increase the value of your statistic: *increase* the between-groups variability or

Between-groups variability Variability in DV scores that is due to the effects of the IV.

Error variability Variability in DV scores that is due to factors other than the IV, such as individual differences, measurement error, and extraneous variation (also known as *within-groups variability*).

decrease the error variability. (Increasing between-groups variability is a function of your IV [see Chapter 6]; we will not discuss that option here.)



Can you figure out why using a correlated-groups design can reduce error variability?

Earlier in this section we listed individual differences as one source of error variability. Correlated-groups designs help reduce this source of error. If, in our treatment groups, we use the same participants or participants who share some important characteristic, either naturally or through matching, those participants will exhibit smaller individual differences between the groups than will randomly chosen participants. Imagine how dissimilar to you another participant could be if we chose that person at random. Imagine how similar to you another participant would be if that person were related to you, had the same intelligence as you, or (in the most obvious situation) were you! If we use a correlated design, then, error variability owing to individual differences should decrease, our statistic should increase, and we should have a greater chance of finding a significant difference as a result of our IV.



Why did we use the hedge word “should” three times in the preceding sentence?

Remember, when we discussed matched pairs earlier, we said that matching on an irrelevant variable could actually hurt your chances of finding a significant difference. If you match on an irrelevant variable, the between-groups differences do not decrease. If the between-groups differences do not decrease, your error variability is the same as if you had used an independent-groups design, which results in identical statistical test results. When we use a statistical test for a correlated-groups design, we must use a larger critical t value than we would have if we conducted the same experiment with an independent-groups design. (The statistical reason for this difference is that we give up some **degrees of freedom** in the correlated-groups design relative to an independent-groups design.) In the two-correlated-groups situation the degrees of freedom are equal to $N - 1$, where N represents the number of *pairs* of participants. In the two-independent-groups situation the degrees of freedom are $N - 2$, where N represents the *total* number of participants.

Degrees of freedom

The ability of a number in a specified set to assume any value.



Suppose you ran an experiment with 10 participants in each group. How many degrees of freedom would you have if this were an independent-groups design? A correlated-groups design?

Did you determine 18 *df* for the independent-groups design and 9 *df* for the correlated-groups design? Using the *t* table in the back of the book (see Table A-1), you will see that the critical *t* value at the .05 level with 18 *df* is 2.101, whereas it is 2.262 with 9 *df*.

These critical *t* values make it seem that it would be easier to reject the null hypothesis in the independent-samples situation (critical *t* = 2.101) than in the correlated-groups situation (critical *t* = 2.262). Yet we said earlier that the correlated-groups situation could benefit us statistically. What's going on?

The preceding numbers *do* support the first sentence of the previous paragraph—it *would* be easier to find a *t* of 2.101 than of 2.262. You must remember, however, that a correlated-groups design should reduce the error variability and result in a larger statistic. Typically, the statistic is increased more than enough to make up for the lost degrees of freedom. Remember that this reasoning is based on the assumption that you have matched on a relevant variable. Matching on an irrelevant variable does not reduce the error variability and will not increase the statistic—in which case, the lost degrees of freedom actually hurt your chances of finding significance. We will show you an actual statistical example of this point at the end of the statistical interpretation section later in this chapter.

Advantages of Independent-Groups Designs The chief advantage of independent-groups designs is their simplicity. Once you have planned your experiment, choosing your participants is quite easy—you merely get a large number of participants and randomly assign them to groups. You don't have to worry about measuring your participants on some variable and then matching them; you don't have to worry about whether each participant can serve in all conditions of your experiment; you don't have to worry about establishing or determining a relationship between your participants—these concerns are relevant only to correlated-groups designs.

Does the statistical advantage of correlated-groups designs render independent-groups designs useless? We cannot argue about the statistical advantage—it is real. However, as you can tell by reviewing the critical *t* values mentioned earlier, the advantage is not overwhelming. As the number of experimental participants increases, the difference becomes smaller and smaller. For example, the significant *t* value with 60 *df* is 2.00, and with 30 *df* it is only 2.04. If you expect your IV to have a powerful effect, then the statistical advantage of a correlated-groups design will be lessened.

One final point should be made in favor of independent-groups designs. Remember that, in some situations, it is simply impossible to use a correlated-groups design. Some circumstances do not allow repeated measures (as we pointed out earlier in the chapter), some participant variables cannot be matched, and some participants cannot be related in any way to other participants.

So what is the elusive conclusion? As you might guess, there is no simple, all-purpose answer. A correlated-groups design provides you with additional control and a greater chance of finding statistical significance. On the other hand, independent-groups designs are simpler to set up and conduct and can overcome the statistical advantages of correlated-groups designs if you use large samples. If you have large numbers of participants and expect your IV to have a large effect, you are quite safe with an independent-groups design. Alternatively, if you have only a small number of experimental participants and you expect your IV to have a small effect, the advantages of a correlated-groups design would be important to you. For all those in-between cases you must weigh the alternatives and choose the type of design that seems to have the greater advantage.

Variations on the Two-Group Design

To this point we have described the two-group design as if all two-group designs were identical. This is not the case. Let's look at two variations on this design.

Comparing Different Amounts of an IV Earlier we said that the most common use of two-group designs was to compare a group of participants receiving the IV (experimental group) to a group that does not receive the IV (control group). Although this is the most common type of two-group design, it is not the only type. The presence-absence manipulation of an IV allows you to determine whether the IV has an effect. For example, McClellan and Woods (2001) were able to determine that customers with a hearing disability received help from salesclerks more slowly than customers without such a disability; Wells (2001) found that her students did react to a mental arithmetic task in a stressful manner. An analogous situation to using the presence-absence manipulation in detective work is trying to sort out the clues that relate to the guilt or innocence of a single suspect.

A presence-absence IV manipulation does not, however, allow you to determine the precise effects of the IV. McClellan and Woods did *not* discover whether having salesclerks wait on deaf customers caused a large or small delay in offering help, only that the clerks responded more slowly. Wells did not determine how stressful mental arithmetic was compared to other tasks, only that it did increase stress.

Typically, after we determine that a particular IV has an effect, we would like to have more specific information about that effect. Can we produce the effect with more (or less) of the IV? Will a different IV produce a stronger (or weaker) effect? What is the optimum amount (or type) of the IV? These are just a few of the possible questions that remain after determining that the IV had an effect. Thus, we can follow up on our IV presence-absence experiment with a new two-group experiment that compares different *amounts* or *types* of the IV to determine their effectiveness. Similarly, a detective may have two suspects and be faced with the task of sorting out the evidence to decide which suspect is more likely the guilty party.

Some IVs simply cannot be contrasted through presence-absence manipulations. For example, Erin Vaughn (2002) of Ouachita Baptist University in Arkadelphia, Arkansas, was interested in the effects of stereotyping on "person perception." In her experiment Vaughn wanted to assess the effects of the tendency to stereotype people on the basis of physical attractiveness (IV) on pairing pictures of men and women into dating couples (DV). This question would make little sense in an IV presence-absence situation: People could show more or less of a tendency to stereotype, but not zero or total tendency. Vaughn therefore compared differing amounts of this IV. Based on participants' answers to a survey, she formed "high tendency to stereotype" and "low tendency to stereotype" groups. When she compared these two groups' pairings of men and women, she found that people with a high tendency to stereotype created more pairings in which men and women were similar in physical attractiveness. Thus, the tendency to stereotype affected the way that participants believed that dating couples "paired up" in Vaughn's experiment.

The key point to notice when we conduct an experiment contrasting different amounts of our IV is that we no longer have a true control group. In other words, there is no group that receives a zero amount of the IV. Again, we are not trying to determine whether the IV has an effect—we already know that it does. We are merely attempting to find a difference between differing types or amounts of our IV.

Dealing with Measured IVs To this point, when we have mentioned IVs in this text, we have emphasized that they are the factors that the experimenter *directly manipulates*.

Technically, this statement is correct only for a **true experiment**. In a true experiment the experimenter has total control over the IV and can assign participants to IV conditions. In other words, the experimenter can manipulate the IV. McClellan and Woods (2001) were able to assign their clerks to either the deaf-customer group or the hearing-customer group. If you wish to assess the effects of two different reading programs on teaching children how to read, you can assign nonreading children to either program.

As you saw in Chapter 6, there are many IVs (participant IVs) that psychologists wish to study but cannot directly manipulate; we measure them instead. For example, Lindsey Smith, a student at Presbyterian College in Clinton, South Carolina, and Marion Gaines, her faculty sponsor, examined performance differences on a perceptual task between college students with attention deficit hyperactivity disorder (ADHD) and students without ADHD. Because they were not able to directly manipulate the ADHD status of their participants (of course), Smith and Gaines (2005) conducted **ex post facto research** (see Chapter 4). In the context of the present chapter they used a two-group design. The participants completed a backward masking perceptual task in which they viewed stimuli on a computer screen that could be masked with Xs to make the perception more difficult. Smith and Gaines found that the participants with ADHD performed more poorly on the perceptual task than the control group. Based on what we know about ADHD, Smith and Gaines's finding may not be surprising. We can even develop hypotheses about why the students with ADHD tended to perform more poorly on the perceptual task; however, we must be cautious in our interpretation. Although we do know from Smith and Gaines's research that participants without ADHD performed better than participants with ADHD, we are not certain why this difference exists. In other words, there could be other differences in the two groups of participants than simply the ADHD. Because Smith and Gaines did not (*could not*) assign participants to groups randomly (the definition of ex post facto research), they could not be certain that ADHD was the only difference between the groups.

The drawback of ex post facto research is certainly a serious one. Conducting an experiment without being able to draw a cause-and-effect conclusion is limiting. Why would we want to conduct ex post facto research if we cannot draw definitive conclusions from it? As we mentioned earlier, some of the most interesting psychological variables do not lend themselves to any type of research other than ex post facto. If you wish to study the genesis of female–male differences, you have no option other than conducting ex post facto studies. Also, as psychologists continue to conduct ex post facto research, they do make progress. Attempting to specify the determinants of intelligence involves ex post facto research—surely you remember the famous heredity-versus-environment debate over IQ. What we *think* we know today is that both factors affect IQ: Psychologists believe that heredity sets the limits of your possible IQ (i.e., your possible minimum and maximum IQs) and that your environment determines where you fall within that range (Weinberg, 1989). Thus, it seems clear that we should not abandon ex post facto research despite its major drawback. We must, however, remember to be extremely cautious in drawing conclusions from ex post facto studies. Detectives, of course, are always faced with ex post facto evidence—it is impossible to manipulate the variables *after* a crime has been committed.

True experiment An experiment in which the experimenter directly manipulates the IV.

Ex post facto research A research approach in which the experimenter cannot directly manipulate the IV but can only classify, categorize, or measure the IV because it is predetermined in the participants (e.g., IV = sex).

■ REVIEW SUMMARY

1. **Correlated-groups designs** provide more control because they guarantee equality of the two groups.
2. Correlated-groups designs generally reduce **error variability** and are more likely to achieve statistically significant results.
3. An advantage of **independent-groups designs** is that they are simple to conduct. With large numbers of research participants, they are also strong designs.
4. Researchers often use two-group designs to compare different amounts (or types) of IVs.
5. We cannot manipulate some IVs, so we must resort to measuring them and conducting **ex post facto research**, which cannot demonstrate cause-and-effect relations.

■ Check Your Progress

1. Why is it important that our two groups be equal before the experiment begins?
2. The variability in DV scores that can be attributed to our experimental treatments is called _____; variability from other sources is labeled _____.
3. Which three factors cause variation in DV scores (error variability)?
 - a. nonrandom assignment, random assignment, and mixed assignment
 - b. individual differences, measurement errors, and extraneous variables
 - c. placebo effects, measurement errors, and the IV
 - d. variance, standard deviation, and the mean
4. Compare and contrast the advantages and disadvantages of independent-groups and correlated-groups designs.
5. Other than the consent form and mental arithmetic IV examples given in the chapter, give two examples of IVs for which you might wish to compare differing amounts.
6. List three examples of IVs not in the text that you would have to study with ex post facto experiments.

Statistical Analysis: What Do Your Data Show?

After you have used your experimental design to conduct an experiment and gather data, you are ready to use your statistical tools to analyze the data. Let's pause for a moment to understand how your experimental design and statistical tests are integrated.

The Relation Between Experimental Design and Statistics

At the beginning of this chapter we compared experimental design to a blueprint and pointed out that you needed a design to know where you were headed. When you carefully plan your experiment and choose the correct experimental design, you also accomplish

another big step. Selecting the appropriate experimental design determines the particular statistical test you will use to analyze your data. Because experimental design and statistics are intimately linked, you should determine your experimental design *before* you begin collecting data to ensure there will be an appropriate statistical test you can use to analyze your data. Remember, you don't want to be a professor's classroom example of a student who conducted research project only to find out there was no way to analyze the data!

Analyzing Two-Group Designs

In this chapter we have looked at one-IV, two-group designs. You may remember from your statistics course, as well as from Chapter 9, that this type of experimental design requires a *t* test to analyze the resulting data (assuming you have interval- or ratio-level data). You may also remember learning about two different types of *t* tests in your statistics class. For a two-independent-groups design you would use a *t* test for independent samples (also known as an independent *t* test) to analyze your data. For a two-correlated-groups design you would analyze your data with a *t* test for correlated samples (also called a dependent *t* test, a within-groups *t* test, or a paired *t* test).

Let's make certain that the relation between experimental design and statistics is clear. A *t* test is indicated as the appropriate statistical test because you conducted an experiment with one IV that has two levels (treatment conditions). The decision of *which t* test to use is based on how you assigned your participants to their groups. If you used random assignment, then you will use the *t* test for independent samples. If you used repeated measures, matched pairs, or natural pairs, then you would use the *t* test for correlated samples.

Calculating Your Statistics

In Chapter 9 we provided the computer analysis of a *t* test. The research example involved a comparison of how long it took salespeople to wait on customers who were dressed in sloppy or dressy clothes. In this chapter we will examine those data more completely. To help set the stage for the remainder of the chapter, let's review some details of the hypothetical experiment behind the data. We wondered whether the clothes students wore actually make any difference in how quickly salesclerks would wait on them. We collected data from 16 different clerks, randomly assigning 8 to wait on customers wearing dressy clothes and 8 to wait on customers wearing sloppy clothes.



Which statistical test would you use to analyze data from this experiment and why?

The simplest way to answer these questions is to use our chart in Figure 10-1. There is only one IV: the type of clothing worn. That IV has two levels: dressy and sloppy. We randomly assigned the participants to their groups. Thus, this design represents a two-independent-groups design, and you should analyze it with an independent *t* test.

Interpretation: Making Sense of Your Statistics

We hope that your statistics instructor taught you this important lesson about statistics: Statistics are not something to fear and avoid; they are a tool to help you understand the data garnered from your experiment. Because of today's focus on computerized statistical analyses, calculating statistics is becoming secondary to interpreting them. Just as having a sewing machine is useless if you don't know how to operate it, statistics are useless if you don't know how to interpret them. Likewise, detectives must learn the skills necessary to interpret the reports they receive from the police scientific labs. We will focus on two types of interpretation in this section: interpreting computer statistical output and translating statistical information into experimental outcomes.

Interpreting Computer Statistical Output

There may be hundreds of computer packages available for analyzing data. Thus, it would be impossible (and inefficient) to show output from every different package and teach you how to interpret each one. Remember that we will show you generic computer statistical output and present interpretations of those analyses. We believe that the similarity among statistical packages will allow you to generalize from our examples to the specific package that you may use. (Computerized statistical packages vary widely in the number of decimal places they report for statistical results. To be consistent with APA format, we will round the computerized output and use only two decimal places in the text.)

The *t* Test for Independent Samples Let's return to our statistical example from Chapter 9. Remember, we randomly assigned clerks to one of two groups: a sloppily dressed group of customers or a well-dressed group. We sent the customers to stores and obtained the time-to-service scores you saw in Chapter 9. If we analyzed these data using a computer package, what might the output look like? We presented an abbreviated version of the output in Chapter 9 for simplicity's sake; a more complete printout appears in Table 10-1.

TABLE 10-1 Computer Output for *t* Test for Independent Groups

GROUP	N	Mean	SD	Standard Error
GROUP 1	8	48.38	10.113	3.575
GROUP 2	8	63.25	12.544	4.435
<i>F</i> _{max test} = 1.634 p = 0.222				
Equal Variances Assumed				
<i>t</i> = 2.61 df = 14 p = 0.021 Cohen's <i>d</i> = 0.92				
Equal Variances Not Assumed				
<i>t</i> = 2.61 df = 13.4 p = 0.021 Cohen's <i>d</i> = 0.92				

We usually examine the descriptive statistics first. The descriptive statistics are printed at the top of the printout. We see that GROUP 1 (defined at the top of the printout as "Dressy Clothing") had 8 cases, a mean salesperson response time of 48.38 seconds, a standard deviation of 10.11 seconds, and a standard error of 3.58 seconds. GROUP 2 (the "Sloppy Clothing" group) had 8 cases, a mean salesperson response time of 63.25 seconds, a standard deviation of 12.54 seconds, and a standard error of 4.44 seconds. Be cautious at this point—an old saying we learned regarding computers is "garbage in, garbage out." In other words, if you enter incorrect numbers into a computer, you will get incorrect numbers out of the computer. You should always verify any numbers you enter and, as much as possible, double-check the output. "Wait a minute," you may be saying. "What's the use of using a computer if I have to check its work?" We're not suggesting that you check up on the computer but that you check up on yourself! For example, suppose the computer information for GROUP 1 or GROUP 2 showed the number of cases to be seven. You would know that the computer didn't read one number—perhaps you entered only seven scores, or perhaps you mislabeled one score. With only eight scores, it is simple enough to calculate the mean for each group yourself. Why should you do that? If you find the same mean that the computer displays, you can be reasonably certain that you entered the data correctly and, therefore, can go on to interpret your statistics.

The second set of statistics provided contains only two statistical values: F and p . These values represent the results of a test known as F_{\max} , a statistic used to test the assumption of **homogeneity of variance** for the two groups (Kirk, 1968). Homogeneity of variance simply means that the variability of the scores of the two groups is similar. To use a t test, we must assume that the variances are similar. In this particular example our assumption is justified because the probability of chance for the F value is .22, well above the standard .05 cutoff. Because we have homogeneity of variance, we will use the third block of information ("Equal Variances Assumed") to interpret our test.

In the second set of information, if our p value were less than .05, we would have found **heterogeneity of variance**, meaning that the variability of the scores of the two groups was *not* comparable. Thus, we would be violating a mathematical assumption for using the t test. Fortunately, statisticians have developed a procedure that allows us to interpret our statistics despite heterogeneity. In such a case we would use the fourth block of information ("Equal Variances Not Assumed") rather than the third block. If the variances of the two groups are equivalent, we can pool or combine those estimates; however, if the variances are not equivalent, we must keep them separate. Again, in our current example, because the F_{\max} statistic is not significant ($p = .22$), we will use the statistical results under the "Equal Variances Assumed" heading.

Generally speaking, t tests are **robust** with regard to the assumption of homogeneity (Kirk, 1968). A robust test is one that can tolerate violations of its assumptions and still provide accurate answers. Kirk noted that the t test is so robust that the homogeneity assumption is often not even tested. The statistics package you use, therefore, may not provide information about the F_{\max} statistic (and thus probably will not give you equal and unequal variance estimates).

Homogeneity of variance

The assumption that the variances are equal for the two (or more) groups you plan to compare statistically.

Heterogeneity of variance

Occurs when we do not have homogeneity of variance; this means that our two (or more) groups' variances are not equivalent.

Robust Refers to a statistical test that can tolerate violation of its assumptions (e.g., homogeneity of variances) and still yield valid results.

By looking at the third set of information, we find that our t value (calculated by the computer) is 2.61. We have 14 degrees of freedom ($N_1 + N_2 - 2$). Rather than having to locate these values in a t table to determine significance, we can use the significance level provided by the computer: The probability (two-tail) is .021. Thus, the probability that two means as different as these could have come from the same population by chance is less than 3 in 100. This probability is less than the magical .05 cutoff, so we conclude that these two means are significantly different (i.e., the difference between them is *not* due to chance).

Some statistical packages may not automatically print the degrees of freedom for you, so it is important to remember how to calculate df . Also, some programs may not provide the probability of your result as part of the printout; then you would have to make this determination yourself. In such a case you would use the t table (Appendix A, Table A-3). In this case you would find that the probability of the t we found is less than .05. (Computer output typically provides exact p values [.021 in this case], whereas statistical tables simply allow you to compare your result to standard p values such as .05 or .01.

In addition, the computer output shows that Cohen's d is 0.92. We learned in Chapter 9 that a d of 0.8 or larger is considered a large effect size. This information helps to confirm that customers' attire plays a major role in determining salesclerks' speed of helping. This decision completes the process of interpreting the computer output. Our next task is to describe our statistical information in terms of the experiment we conducted.

Translating Statistics Into Words Think back to the logic of an experiment: We start an experiment with two equal groups and treat them identically (for control purposes) with one exception (our IV, or type of dress); we measure the two groups (on our DV, or time to provide service) in order to compare them. At this point, based on our statistical analyses, we know that we have a significant difference (i.e., not due to chance) between our two means. If two equal groups began the experiment and they are now *unequal*, to what can we attribute that difference? If our controls have been adequate, our only choice is to assume that the difference between the groups is due to the IV.

Looking at our example, we have decided that the groups of students dressed in two different types of clothing received help from clerks in different amounts of time. Many students stop at this point, thinking that they have drawn a complete conclusion from their experiment.



Why would this conclusion be incomplete? Can you develop a complete conclusion before proceeding?

Saying that students who are dressed differently get waited on in different amounts of time is an incomplete conclusion because it specifies only a difference, not the *direction* of that difference. Whenever we compare treatments and find a difference, we want to know which group has performed at a better or higher level. In a two-group experiment this interpretation is quite simple. Because we have only two groups and we have concluded that they differed significantly, we can further conclude that the group with the higher mean score has outscored the group with the lower mean score (remember that high scores do not always indicate superior performance, as in this case).

To interpret fully the results from this experiment, we examine our descriptive statistics and find that the salespeople waiting on students dressed in sloppy clothes had a mean response time of 63.25 seconds, whereas the salespeople waiting on the well-dressed students averaged 48.38 seconds. Thus, we can conclude that the salesclerks waiting on nicely dressed customers responded more quickly than those waiting on sloppily dressed customers. Notice that this statement includes both the notion of a difference *and* the direction of that difference.

When we draw conclusions from our research, we want to communicate those results clearly and concisely in our experimental report. To accomplish these two objectives, we use both words and numbers in our communication. This communication pattern is part of the APA style for preparing research reports, which we will consider in Chapter 14 (APA, 2001). We will introduce the form for statistical results here. Bear in mind that you are trying to communicate—to tell what you found in words and provide statistical information to support those words. For example, if you were writing an interpretation of the results from our sample experiment, you might write something like the following:

Salesclerks who waited on well-dressed customers ($M = 48.38, SD = 10.11$) took significantly less time, $t(14) = 2.61, p = .021$, to respond to customers than salespeople who waited on customers dressed in sloppy clothing ($M = 63.25, SD = 12.54$). The effect size, estimated with Cohen's d , was .92.

Notice that the words alone give a clear account of the findings—a person who has never taken a statistics course could understand this conclusion. The inferential statistics regarding the test findings support the conclusion. The descriptive statistics (M = mean, SD = standard deviation) given for each group allow the reader to see how the groups actually performed and to see how variable the data were. This standard format allows us to communicate our statistical results clearly and concisely.

The t Test for Correlated Samples Remember that we have covered two different two-group designs in this chapter. Now we will examine the computer output for analysis of the two-correlated-groups design. Our experiment concerning the salespeople was an example of the two-independent-groups design, which would *not* require a t test for correlated samples.



How could you modify this experiment so that it used correlated groups rather than independent groups?

You should remember that there are three methods for creating correlated groups: matched pairs, repeated measures, and natural pairs. If your modified experiment used one of these techniques, you made a correct change. As an example, let's assume that we were worried about the difference between salesclerks confounding our experiment. To better equate the clerks in our two groups, we decide to use the repeated-measures approach. We decide to measure each salesclerk's time to respond to two customers: once for a dressed-up customer and once for a sloppily dressed customer. Before beginning our experiment, we know that the salespeople are identical for the two groups, thus removing individual differences as a potential confounding variable.

Next, we conduct our experiment. We measure the response time of each of the eight clerks waiting on both types of customers (based on dress). Given this hypothetical example, the

TABLE 10-2 Computer Output for *t* Test for Correlated Groups

GROUP 1 = Dressy clothing GROUP 2 = Sloppy clothing Variable = Salesclerks' response time				
	N	Mean	SD	Standard Error
GROUP 1	8	48.38	10.113	3.575
GROUP 2	8	63.25	12.544	4.435
Mean difference =	14.875	SD = 7.699	Std Error = 2.722	
Corr. =	0.790	p = 0.020		
<i>t</i> =	5.465	df = 7	p = 0.001	Cohen's <i>d</i> = 1.93

scores from Chapter 9 would now represent repeated-measures time scores rather than independent scores. After analyzing the data with our computer package, we find the output in Table 10-2. (Please note that it is *not* legitimate to analyze the same data with two different statistical tests. We are doing so in this chapter merely for example's sake. If you tested real-world data multiple times, you would increase the probability of making a Type I error; see Chapter 9.)

Look at Table 10-2. Again, we first look for the descriptive statistics and find them at the top of the printout. Of course, because we used the same data, we have the same descriptive statistics as for the independent-samples test. Salesclerks waiting on the students wearing sloppy clothing responded in an average of 63.25 seconds, with a standard deviation of 12.54 and a standard error of 4.44. The students who wore dressy clothes received help in 48.38 seconds, with a standard deviation of 10.11 and a standard error of 3.58. Remember that there are 8 pairs of scores (representing the 8 clerks) rather than 16 individual scores. This difference between the two *t* tests will be important when we consider the degrees of freedom.

The second block of information shows us the size of the difference between the two means, as well as its standard deviation and standard error. (Researchers rarely use this information, so it may not appear in your computer output.) The third block gives you some information about the relation between the pairs of participants (or the same participant for repeated measures). Here you can determine whether the paired scores were correlated. Remember that we want them to be correlated so that we will gain the additional statistical control made available by using the correlated-groups design. As you can see in this example, the scores were highly **positively correlated** (see Chapters 4 and 9). In our example, this result implies that if a salesclerk waited on one student quickly, he or she tended also to wait on the other student quickly.

In the fourth block we find the results of our inferential test. We obtained a *t* value of 5.47 with 7 degrees of freedom.

Positive correlation As scores on one variable increase, scores on the second variable also increase.



We have 16 data points (8 clerks measured twice each) in our experiment but only 7 degrees of freedom. In our earlier example, we had 16 participants and 14 degrees of freedom. What is the difference in this case?

You should remember that our degrees of freedom for correlated-samples cases are equal to the *number of pairs of participants minus 1*. If this is fuzzy in your memory, refer to Statistical Issues earlier in the chapter.

The computer tells us that the probability of a *t* of 5.47 with 7 *df* is .001. With such a low probability of chance for our results, we would conclude that there is a significant difference between the clerks' response times to differently dressed students. In other words, we believe that it is highly unlikely that the difference between our groups could have occurred by chance and that, instead, the difference must be due to our IV. The effect size information, Cohen's *d*, provides ample support for our conclusion as *d* is 1.93. Remember that 0.8 represents a large effect size; therefore, the effect of the IV is quite substantial in this analysis.

Translating Statistics Into Words Our experimental logic is exactly the same for this experiment as it was for the independent-samples case. The only difference is that with our matched participants, we are more certain that the two groups are equal before the experiment begins. We still treat our groups equally (control) with the one exception (our IV) and measure their performance (our DV) so that we can compare them statistically.

To translate our statistics into words, it is important to say more than the fact that we found a significant difference. We must know what form or direction that significant difference takes. With the *t* test for correlated samples, we are again comparing two groups, so it is a simple matter of looking at the group means to determine which group outperformed the other. Of course, because we are using the same data, our results are identical: The sloppily dressed students received help in a mean of 63.25 seconds compared to 48.38 seconds for the well-dressed students.



How would you write the results of this experiment in words and numbers for your experimental report?

Did you find yourself flipping back in the book to look at our earlier conclusion? If so, that's a good strategy because this conclusion should be quite similar to the earlier conclusion. In fact, you could almost copy the earlier conclusion as long as you made several important changes. Did you catch those changes? Here's an adaptation of our earlier conclusion:

Salespeople who waited on well-dressed customers ($M = 48.38$, $SD = 10.11$) took significantly less time, $t(7) = 5.47$, $p = .001$, to respond to the customers than when they waited on customers dressed in sloppy clothes ($M = 63.25$, $SD = 12.54$). The effect size, estimated with Cohen's *d*, was 1.93.

As you can see, four numbers in the sentences changed: We had fewer degrees of freedom, our *t* value was larger, our probability of chance was lower, and our effect size was much larger. In this *purely hypothetical* example in which we analyzed the same data twice (a clear violation of assumptions if you were to do this in the real world), you can see the advantage of correlated-groups designs. Although we lost degrees of freedom compared to the independent-samples case presented earlier in the chapter, the probability that our results were due to chance

actually decreased and our effect size increased dramatically. Again, we gained these advantages because our matching of the participants decreased the variability in the data.

You can see a vivid illustration of what we gained through matching by examining Table 10-3. In this example we have used the same data from Chapter 9 but shuffled the scores for the second group before we ran a *t* test for correlated samples. Such a situation would occur if you matched your participants on an irrelevant variable (remember that we mentioned this possibility earlier in the chapter). As you can see by comparing Tables 10-3 and 10-2, the descriptive statistics remained the same because the scores in each group did not change. However, in Table 10-3 the correlation between the two sets of scores is now $-.88$. Because there is a **negative correlation** between our scores, the *t* value is 1.91, even lower than it was in our *t* test for independent groups (see Table 10-1). The marked change comes when we compare the inferential statistics in Tables 10-2 and 10-3. The original analysis showed a *t* of 5.47 with $p = .001$. In contrast, with a negative correlation between the pairs of scores, the new analysis shows a *t* of 1.91 with $p = .097$ and an effect size of 0.68 (the smallest of our three analyses). Thus, these results did not remain significant when the correlation between the participants disappeared. Again, the key point to remember is that when using a correlated-groups design, the groups should actually be positively correlated.

Negative correlation

As scores on one variable increase, scores on the second variable decrease.

The Continuing Research Problem

Research is a cyclical, ongoing process. It would be rare for a psychologist to conduct a single research project and stop at that point because that one project had answered all the questions about the particular topic. Instead, one experiment usually answers some of your questions, does not answer others, and raises new ones for your consideration. As you have studied the work of famous psychologists, you may have noticed that many of them established a research area early in their careers and continued working in that area for the duration of their professional lives. We're not trying to say that the research area you choose as an undergraduate will shape your future as a psychologist—although it could! Rather, we are merely pointing out that one good experiment often leads to another.

TABLE 10-3 Computer Output for *t* Test for Correlated Groups (Shuffled Data)

Group	<i>N</i>	Mean	<i>SD</i>	Standard Error
GROUP 1	8	48.38	10.113	3.575
GROUP 2	8	63.25	12.544	4.435
Mean difference	= 14.88	<i>SD</i> = 21.977	Std Error = 7.770	
Corr.	= $-.880$	<i>p</i> = 0.004		
<i>t</i>	= 1.91	<i>df</i> = 7	<i>p</i> = 0.097	Cohen's <i>d</i> = 0.68

We want to show you how research is an ongoing process as we move through the next two chapters with our continuing research problem. We sketched out a research problem in Chapter 9 (comparing how customers' style of dress affects salespeople's performance) and asked you to help us solve the problem through experimentation. We will continue to examine this problem throughout the next two chapters so that you can see how different questions we ask about the same problem may require different research designs. This research problem is purely hypothetical, but it has an applied slant to it. We hope the continuing research problem helps you see how a single question can be asked in many different ways and that a single question often leads to many new questions.

To make certain you understood the logical series of steps we took in choosing a design, let's review those steps, paying particular attention to the experimental design questions shown in Figure 10-1:

1. After reviewing relevant research literature, we chose our IV (style of dress) and our DV (salesclerk response time).
2. Because we were conducting a preliminary investigation into the effects of clothing on salesclerks' reactions, we decided to test only one IV (the style of dress).
3. Because we wanted to determine only whether clothing style can affect the performance of salespeople, we chose to use only two levels of the IV (dressy clothing vs. sloppy clothing).
- 4a. If we have a large number of participants available, then we can use random assignment, which yields independent groups. In this case we would use the two-independent-groups design and analyze the data with a *t* test for independent groups.
- 4b. If we expect to have a small number of participants and must exert the maximum degree of control, we choose to use a design with repeated measures or matched groups, thus resulting in correlated groups. Therefore, we would use a two-correlated-groups design for the experiment and analyze the data with a *t* test for correlated groups.
5. We concluded that salespeople responded more quickly to customers in dressy clothes than to customers dressed in sloppy clothes.

■ REVIEW SUMMARY

1. The statistical test you use for analyzing your experimental data is related to the experimental design you choose.
2. When you have one IV with two groups and use randomly assigned research participants, the appropriate statistical test is the ***t* test for independent samples**.
3. When you have one IV with two groups and use matched pairs, natural pairs, or repeated measures with your participants, the appropriate statistical test is the ***t* test for correlated samples**.
4. Computer printouts of statistics typically give descriptive statistics (including means and standard deviations) and inferential statistics.
5. To communicate the statistical results of an experiment, we use APA format for clarity and conciseness.
6. Research is a cyclical, ongoing process. Most experimental questions can be tested with different designs.

Check Your Progress

1. Which statistical test would you use if you compared the stereotyping of a group of female executives to a group of male executives? Explain your reasoning.
2. Which statistical test would you use if you compared the stereotyping of a group of male executives before and after an antidiscrimination bill passed through Congress? Explain your reasoning.
3. Compared to the *t* test for independent groups, the *t* test for correlated samples has _____ degrees of freedom.
 - a. fewer
 - b. more
 - c. exactly the same number of
 - d. none of these; it is impossible to compare degrees of freedom between statistical tests
4. What information do we usually look for first on a computer printout? Why?
5. If the variability of our two groups is similar, we have _____; if the variability of the groups is dissimilar, we have _____.
6. When we write a report of our experimental results, we explain the results in _____ and _____.
7. Interpret the following statistics:
Group A ($M = 75$); Group B ($M = 70$); $t(14) = 2.53, p < 0.05$
8. Why do we describe research as a cyclical, ongoing process? Give an example of how this cycling might take place.

Key Terms

Experimental design, 203	Between-subjects comparison, 208	Ex post facto research, 219
Principle of parsimony, 204	Confounded experiment, 208	Homogeneity of variance, 223
Independent variable (IV), 204	Correlated assignment, 209	Heterogeneity of variance, 223
Dependent variable (DV), 204	Matched pairs, 210	Robust, 223
Extraneous variables, 206	Repeated measures, 211	Positive correlation, 226
Levels, 206	Natural pairs, 212	Negative correlation, 228
Experimental group, 206	Within-subjects comparison, 212	
Control group, 206	Between-groups variability, 215	
Random assignment, 207	Error variability, 215	
Random selection, 208	Degrees of freedom, 216	
Independent groups, 208	True experiment, 219	

Looking Ahead

In this chapter we have examined the notion of planning an experiment by selecting a research design. In particular we examined the basic building-block designs with one IV and two groups. In the next chapter we will enlarge this basic design by adding more groups to our one IV. This enlarged design will give us the capability to ask more penetrating questions about the effects of our IV and to obtain more specific information about those effects.

Designing, Conducting, Analyzing, and Interpreting Experiments with More Than Two Groups

Experimental Design: Adding to the Basic Building Block

- The Multiple-Group Design
- Comparing the Multiple-Group and Two-Group Designs • Comparing Multiple-Group Designs
- Variations on the Multiple-Group Design

Statistical Analysis: What Do Your Data Show?

- Analyzing Multiple-Group Designs • Planning Your Experiment • Rationale of ANOVA

Interpretation: Making Sense of Your Statistics

- Interpreting Computer Statistical Output

The Continuing Research Problem

Experimental Design: Adding to the Basic Building Block

In Chapter 10 we learned many concepts and principles about **experimental design** that are basic to planning *any* experiment, not merely the basic two-group experiment. When we come to one of those topics in this chapter, we will briefly review it and refer you back to Chapter 10 for the original discussion.

In this chapter we will add to our basic building-block design. Consider our previous analogy: As a child you quickly mastered the beginner's set of Legos or Tinkertoys. You learned to build everything there was to build with that small set and then wanted to go beyond those simple objects to build larger, more exciting creations. To satisfy this desire, you got a larger set of building materials that you could combine with the starter set in order to build more complicated objects. Despite the fact you were using a larger set of materials, the basic principles you learned with your starter set still applied.

Experimental design works in much the same way. Researchers typically want to move beyond two-group designs so that they can ask more complicated, more interesting questions. Fortunately, they don't have to start from scratch—that's why we referred to the two-group design as the basic building-block design in the previous chapter. Every experimental design is based on the two-group design. Although the questions you ask may become more complicated or sophisticated, your experimental design principles will remain constant. In the same way, when they face a more difficult case, detectives continue to use the basic investigative procedures they have learned.

Experimental design

The general plan for selecting participants, assigning participants to experimental conditions, controlling extraneous variables, and gathering data.

It is still appropriate to think of your experimental design as the blueprint for your experiment. We hope the following analogy convinces you of the need for having an experimental design. Although you *might* be able to get by without a blueprint if you're building a doghouse, it is unlikely you would want to build your own house without a blueprint. Think of building a small house as being equivalent to using the two-group design from Chapter 10. If you need a blueprint to build a house, imagine how much more you would need a blueprint to build an apartment building or a skyscraper. We will work toward the skyscrapers of experimental design in Chapter 12.

The Multiple-Group Design

Independent variable (IV)

A stimulus or aspect of the environment that the experimenter directly manipulates to determine its influence on behavior.

Here, we will consider an extension of the two-group design. Turn back to Figure 10-2 (page 207) for just a moment. What would be the next logical step to add to this design so that we could ask (and answer) slightly more complex questions?

How Many IVs? The first question that we ask when considering any experimental design is always the same: "How many **independent variables (IVs)** will I use in my experiment?" (see Figure 11-1). In this chapter we will continue to consider only experiments that use one IV. We should remember that although one-IV

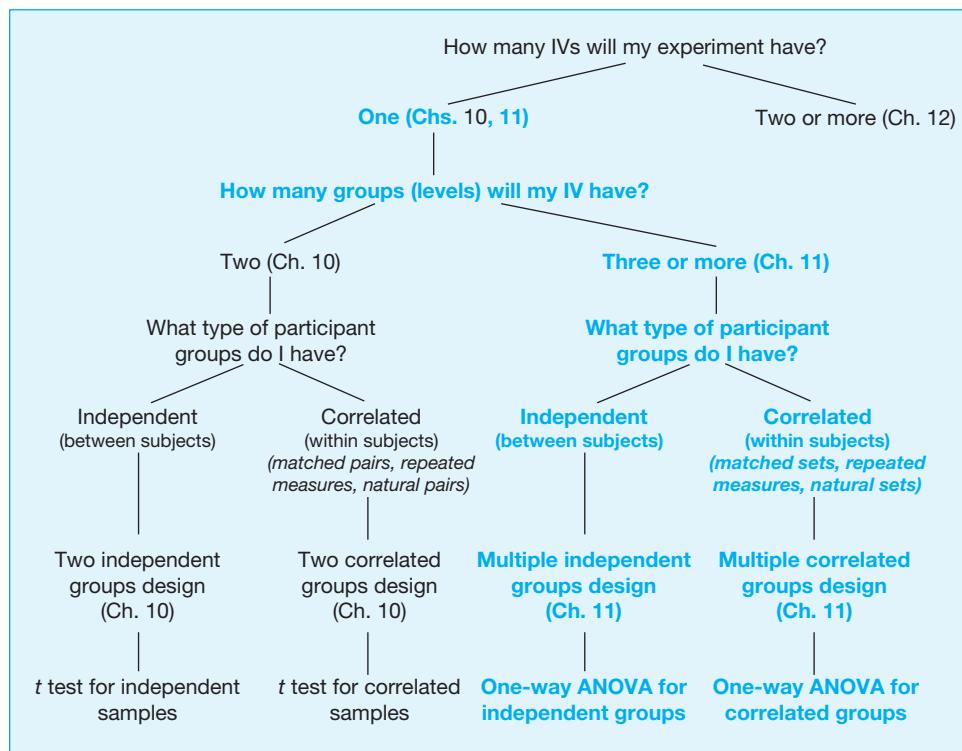


FIGURE 11-1 Experimental Design Questions.

experiments may be simpler than experiments that use multiple IVs (Chapter 12), they are not inferior in any way. Many students who are designing their first research study decide to throw in everything except the kitchen sink as an IV. A well-designed experiment with one IV is vastly preferable to a sloppy experiment with many variables thrown together. Remember the **principle of parsimony** from Chapter 10: If a one-IV experiment can answer your questions, use it and don't complicate things needlessly.

How Many Groups? As soon as we have decided to conduct a one-IV experiment, our second question (see Figure 11-1) revolves around how many groups we will use to test the IV. This question marks the difference between the multiple-group design and the two-group design. As their names imply, a two-group design compares two **levels** of an IV, whereas a multiple-group design compares three or more levels of a single IV. Thus, we could compare three, four, five, or even more differing levels or amounts of an IV. This experimental situation is similar to that experienced by a detective faced with multiple suspects. Instead of merely investigating two people, the detective must conduct simultaneous investigations of more than two individuals.



In Chapter 10 we learned that the most common type of two-group design uses experimental and control groups. How will the multiple-group design differ from that common two-group design?

Actually, there are two answers to that question—if you got either, pat yourself on the back. First, a multiple-group design *can* have a control group. Rather than having a single experimental group and a control group, a multiple-group design with a control group would also have two or more experimental groups. This combination allows us to condense several two-group experiments into one experiment. Instead of conducting a two-group experiment to determine whether your IV has an effect and a second two-group experiment to determine the optimum amount of your IV, you could conduct a multiple-group experiment with a control group and the number of **treatment groups** you would like to assess. Second, a multiple-group design does not have to have a control group. If you already know that your IV has an effect, you can simply compare as many treatment groups as you would like in your multiple-group design.

Let's look at a research example using the multiple-group design. Colleen Sullivan and her faculty advisor, Camille Buckner (2005), of Frostburg State University in Frostburg, Maryland, wanted to determine whether the type of role model affected students' intentions to purchase products. They used parental, peer, celebrity, and no role model conditions. Why does this experiment fit the multiple-group design? First, it has one IV: the type of role model. Second, the IV has more than two levels: It has four, based on the four different role model conditions. Thus, as you can see in Figure 11-1, with one IV and four levels, this experiment requires a multiple-group design. We can draw the block diagram depicted in Figure 11-2 to

Principle of parsimony

The belief that explanations of phenomena and events should remain simple until the simple explanations are no longer valid.

Levels Differing amounts or types of an IV used in an experiment (also known as *treatment conditions*).

Treatment groups

Groups of participants that receive the IV.

INDEPENDENT VARIABLE (TYPE OF ROLE MODEL)

EXPERIMENTAL GROUP 1	EXPERIMENTAL GROUP 2	EXPERIMENTAL GROUP 3	EXPERIMENTAL GROUP 4
Parent	Peer	Celebrity	No role model

FIGURE 11-2 The Multiple-Group Design Used by Sullivan and Buckner (2005).

Source: From "The Influence of Perceived Role Models on College Students' Purchasing Intention and Product-Related Behavior," by C. J. Sullivan and C. E. Buckner, 2005, *Psi Chi Journal of Undergraduate Research*, 10, pp. 66–71.

portray this experimental design. By comparing Figures 10-2 and 11-2, you can easily see how the multiple-group design is an extension of the two-group design.

Three of the groups shown in Figure 11-2 are experimental groups. Does this experiment use a control group? Yes, the *no role model* condition served as a control group. In this experiment, Sullivan and Buckner were interested in the various differences among the four groups based on their differing types of role models. They found that students in the parental role model condition were higher in their intention to buy products than students in the celebrity and no role model conditions; students in the peer role model condition scored between the parental and other two conditions but were not significantly different from any group. In statistical terminology, then, Sullivan and Buckner found results that supported the experimental hypothesis (i.e., that there was a difference between the performance of the groups as a function of role model condition). Of course, support for the experimental hypothesis is not the same as *proving* the experimental hypothesis. Did Sullivan and Buckner *prove* that the type of role model affects students' buying intentionality? No, they merely demonstrated that there was a difference that was unlikely to have occurred by chance for the role model conditions they used and for their groups of participants. What about using different role models? What about using different research participants—children or older adults, for example? Recall that in Chapter 8 we talked about how to generalize our results beyond the specific participants in our experiment.



Suppose you wished to test more than three conditions. Could you use a multiple-group design in such a case? Why or why not? If so, what would happen to the block design in Figure 11-2?

Yes, you could use the multiple-group design if you had four or five role models to assess. In fact, it could be used if there were 10 or 20 role models. The only requirement for using the multiple-group design is an experiment with one IV and more than two groups (see Figure 11-1). Practically speaking, it is rare that multiple-group designs are used with more than four or five groups. If we did use such a design with more than three groups, we would merely extend our block diagram, as shown in Figure 11-3.

Assigning Participants to Groups After we decide to conduct a multiple-group experiment, we must decide about the assignment of research participants to groups (see

INDEPENDENT VARIABLE (TYPE OF ROLE MODEL)				
Expl. Group 1	Expl. Group 2	Expl. Group 3	Expl. Group 4	Expl. Group 5
Role model 1	Role model 2	Role model 3	Role model 4	Role model 5

FIGURE 11-3 Hypothetical Multiple-Group Design With Five Groups.

Figure 11-1). Just as in Chapter 10, we choose between using **independent groups** or **correlated groups**.

Independent Samples (Random Assignment to Groups) Remember that with **random assignment** each participant has an equal chance of being assigned to any group. In their experiment on the effects of role models on buying intention, Sullivan and Buckner (2005) used random assignment when they assigned students to groups: All 69 students had a 1 in 4 chance of being in the parental, peer, celebrity, or no role model group. When we use large groups of participants, random assignment should create groups that are equal on potential extraneous variables such as personality, age, and sex. Recall from Chapter 9 that random assignment allows us to control extraneous variables about which we are unaware. Thus, random assignment serves as an important **control procedure**. For example, we would not want to use role models with only male or female college students. We want to spread the different levels of the IV across all types of participants in order to avoid a **confounded experiment**. Suppose we put all women in the celebrity role model condition and all men in the parental role model condition. When we tabulated our results, we would not be able to draw a clear conclusion about the effects of type of role model because role model was confounded with participant sex. In other words, we couldn't be sure whether a significant difference between groups was caused by the role model difference between the groups or the sex difference between the groups.

Random assignment results in participants who have no relation to participants in other groups; in other words, the groups are independent of each other. We are interested in comparing differences between the various independent groups. As shown in Figure 11-1, when we use random assignment in this design, we end up with a multiple-independent-groups design.

Correlated Samples (Nonrandom Assignment to Groups) In the multiple-group design, we have the same concern about random assignment that we did with the two-group design: What if the random assignment does not work and we begin our experiment with unequal groups? We know that random assignment *should* create equal groups but also that it is most likely to work in the long run—that is, when we have many participants. If we have few participants or if we expect only small differences owing to our IV, we may want more control than random assignment affords us. In such situations, we often resort to using nonrandom methods of assigning participants to groups and thus end up with

Independent groups
Groups of participants formed by random assignment.

Correlated groups
Groups of participants formed by matching, natural pairs, or repeated measures.

Random assignment A method of assigning research participants to groups so that each participant has an equal chance of being in any group.

Control procedure One of several steps experimenters take to ensure that potential extraneous variables are controlled, including random assignment, matching, and so on.

Confounded experiment
An experiment in which an extraneous variable varies systematically with the IV, which makes drawing a cause-and-effect relation impossible.

cathy®

by Cathy Guisewite



Cathy © Cathy Guisewite. Reprinted with permission of Universal Press Syndicate. All rights reserved.

Is Cathy using random assignment or random selection in this cartoon? You may want to look back at Chapters 6 and 7.

Because she is tasting chocolate Santas in a nonsystematic way (rather than assigning Santas to groups), Cathy's gluttony illustrates random selection.

correlated groups. Let's examine our three ways of creating correlated groups and see how they differ from the strategies discussed in Chapter 10.

1. Matched Sets. Matched *pairs* are not appropriate for the multiple-group design because we have at least three groups; therefore, we must use matched *sets*. The principle for forming matched sets is the same as that for forming matched pairs. Before our experiment we measure our participants on some variable that will affect their performance on the DV. Then we create sets of participants who are essentially the same on this measured variable, often known as the **matching variable**. The size of the sets will, of course, depend on how many levels our IV has. If our IV has five levels, for example, then each set would have five participants equated on the matching variable. After we create our matched sets, we then randomly assign the participants within each set to the different groups (treatments).

Returning to Sullivan and Buckner's (2005) experiment, suppose we believed that participant sex would be an extraneous variable because we were using an attractive male for the celebrity role model condition. To ensure equality of their groups on sex, Sullivan and Buckner could have used sets of four participants who were matched on sex, with each participant then randomly assigned to one of the four groups. In this manner, they would have assured that the distribution of participant sex was uniform across all three groups. If all four of their groups had the same sex composition (regardless of whether it was 50-50), then participant sex could not be an extraneous variable.

One final caution is in order. We should remember that the potential extraneous variable must actually affect performance on the DV or else we have hurt our chances of finding a significant difference.

2. Repeated Measures. Other than the fact that participants perform in three or more conditions rather than only two, repeated measures in the multiple-group design are identical to repeated measures in the two-group design. When you use repeated measures in the multiple-group design, each participant must take part in *all* the various

Matching variable A potential extraneous variable on which we measure our research participants and from which we form sets of participants who are equal on the variable.

treatment conditions. Thus, for Sullivan and Buckner to have used repeated measures, each student would have to have participated in the parental, peer, celebrity, and no role model conditions.



Can you see any possible flaws in conducting Sullivan and Buckner's experiment using a repeated-measures design? (Hint: Consider some of the practical issues from Chapter 10 about using repeated measures.)

Several problems could occur if we attempted the role modeling experiment with a repeated-measures design. Would you have one student complete the same survey four different times? This approach would not seem to make any sense. The students might become suspicious about the differing role models; they would probably be able to figure out what the IV was. Or they might get bored from completing the survey four times and not take it seriously on the third or fourth administration. What about using four different surveys? Again, this approach could be problematic. Could we logically assume that students gave different responses because of the role model? Not necessarily, because different surveys would have to differ in some way from one other. Students might give different types of answers to different types of surveys. Thus, we might be measuring the response to different surveys rather than to different types of role models! It seems that this experiment simply would not work well as a repeated-measures design. Remember that we mentioned this possible problem in Chapter 10—not all experiments can be conducted using repeated measures.

3. **Natural Sets.** Using natural sets is analogous to using natural pairs except that our sets must include more than two research participants. Using multiple groups takes away our ability to use some interesting natural pairs such as twins or husbands and wives, but other possibilities for using natural sets do exist. For example, many animal researchers use littermates as natural sets, assuming that their shared heredity makes them more similar than randomly selected animals. In a similar fashion, if your research participants were siblings from families with three (or more) children, natural sets would be a possibility. Most natural pairs or sets involve biological relationships.

We create multiple-correlated-groups designs when we use matched sets, repeated measures, or natural sets. The critical distinction is that the participants in these types of groups are related to each other in some way—we are comparing differences *within* groups (or within subjects, to use the old terminology). On the other hand, in independent-groups designs, the participants have no common relationship. Thus, we compare differences *between* differing groups of subjects.

Kimberly Walker and her faculty advisors James Arruda and Keegan Greenier (1999) of Mercer University in Macon, Georgia, conducted an experiment that used a multiple-correlated-groups design. They were interested in measuring how accurately people respond on the Visual Analogue Scale (VAS). The VAS is a self-report scale used for people to indicate internal states such as mood, hunger, or pain. Participants respond on the VAS by making a mark somewhere along a line, typically 100 millimeters (mm) in length, which has marks at the ends designating extreme low and extreme high values of the internal state. Walker et al.

asked participants to make marks at distances of 10, 20, 30, 40, 50, 60, 70, 80, and 90 mm along the line. They calculated participants' errors by subtracting the difference between the actual distance and the marked distance. By measuring the errors over nine distances, Walker et al. could compare data to determine whether errors were more likely to occur at different points along the line.



Why does the Walker, Arruda, and Greenier experiment illustrate a correlated-groups design? Which particular correlated-groups technique did they use? Why do you think they used a correlated-groups design?

In this experiment the IV was distance, measured in mm. Thus, there were nine levels of the IV. They measured each participant on each different distance; therefore, the correlated groups were a result of using repeated measures. Walker et al. probably used a repeated-measures design because it would take an enormous number of participants if they had each participant serve in only one distance group (nine groups of people needed). Also, it would have been an inefficient use of participants to use nine different groups. Finally, by having people participate in all levels of the IV, Walker et al. did not have to worry about the assumption that the groups were equivalent before the experiment began. By using each person as his or her own control across the nine measurements, the question of group equality was answered. The use of repeated measures helped control many subject variables that *might* have affected participants' performance on the VAS—factors such as motivation, spatial ability, and sex.

Walker et al. (1999) found that the participants were more accurate in making marks at the beginning and the end of the VAS line than in the middle of the line. They explained their results as being due to a perceptual phenomenon: the law of visual angle. Furthermore, they recommended adding depth cues to the VAS to increase accuracy on the task.

Comparing the Multiple-Group and Two-Group Designs

As in Chapter 10, we have to make a decision about how to design our potential experiment. Just as in Chapter 10, there are two multiple-group designs from which to choose. Researchers who want to design an experiment with one IV, however, must also choose between multiple-group designs and two-group designs. In the following sections, we will examine the various advantages and disadvantages of these experimental designs. As we warned you in Chapter 10, read carefully—you may be facing this choice yourself in the future!

The multiple-group design is quite similar to the two-group design. As a matter of fact, all you have to do to change your two-group design into a multiple-group design is add another level (or more) to your IV. Given this high degree of similarity, how would we compare these two designs?

In choosing a design for your experiment, your paramount concern is your experimental question. Does your question require only two groups to find an answer, or does it necessitate three or more groups? This question almost seems like a "no-brainer," but it cannot be taken for granted. Following the principle of parsimony from Chapter 10, we want to select the simplest possible design that will answer our question.

In Chapter 10 we provided you with an ideal situation for a two-group design—an experiment in which we merely wished to determine whether our IV has an effect. Often such an experiment is not necessary because that information already exists in the literature. You should never conduct an experiment to determine whether a particular IV has an effect without first conducting a thorough literature search (see Chapter 2). If you find no answer in a library search, then you should consider conducting a two-group (presence–absence) study. If, however, you find the answer to that basic question and wish to go farther, a multiple-group design might be appropriate.

After these considerations, what do you do when you face a situation in which either a two-group design or a multiple-group design is appropriate? Although this answer may sound odd, you should think about your (future) results. What will they look like? What will they mean? Most critically, what will the addition of any group(s) beyond the basic two tell you? If the information that you expect to find by adding a group or groups is important and meaningful, then by all means add the groups. If, however, you're not really certain what you might learn by adding to the two groups, then you may be merely complicating your experiment needlessly.

Think back to the two student examples cited in this chapter. Did the researchers learn important information by adding an extra group to their two groups? Sullivan and Buckner (2005) found that intent to buy a product differed with four different types of role models (parent, peer, celebrity, and no model). Some people might wish to examine more levels in this experiment rather than fewer. In fact, one of your first thoughts when you read about Sullivan and Buckner's experiment earlier may have been, "I wonder how students would respond to _____ as a role model?" (Fill in the blank with a person or entity whom you believe would be particularly persuasive, e.g., an athlete.) You may believe that this experiment was not a fair test of the question about buying based on role model, especially if you disagree with their role model choices. It appears that Sullivan and Buckner made a wise decision in using a multiple-group design rather than a two-group design. In fact, it might have been more informative had they used an even larger multiple-group design.

Walker et al. (1999) measured people's responses on the VAS to nine different distances. They clearly benefited by using a multiple-group design. If they had simply measured the responses on short or long distances, they would have found little error in the responses.



Suppose Walker et al. (1999) wanted to use a smaller or larger multiple-group design—say measuring people's responses over 5 or 15 distances rather than 9. Would it be possible to use such a small or large multiple-group design?

Could you use a multiple-group design with 5 or 15 groups or measurements? Of course you could. The only limitation would be a practical consideration: Could you secure enough participants to take part (for matched sets or natural sets), or can the participants cope with being measured so many times (for repeated measures)? Because the Walker et al. experiment used repeated measures, our concern there would be for the experimental participants. In this case the participants would be measured either fewer or more times. This experiment, then, would be no problem to run over 5 measurement distances; participants would have fewer responses

to make. The question is whether 5 distances would give the experimenters all the information they needed. On the other hand, using 15 measurements might make the task extremely difficult for the participants to discriminate, and it would require more of the participants' time. Using nine levels of an IV for repeated measures is quite unusual, so it is much more likely that other experimenters would use smaller, rather than larger, experimental designs.

In summary, the multiple-group and two-group designs are quite similar; however, there are important differences between them that you should consider when choosing an experimental design for your research project.

Comparing Multiple-Group Designs

As you might guess, our comparison of the multiple-independent-groups design to the multiple-correlated-groups design is going to be fairly similar to our comparison of the two-group designs in Chapter 10. Practical considerations become somewhat more important, however, in the multiple-group designs, so our conclusions will be somewhat different.

Choosing a Multiple-Group Design Again, your first consideration in choosing an experimental design should always be your experimental question. After you have decided on an experiment with one IV and three or more groups, you must determine whether you should use independent or correlated groups. If only one of those choices is viable, you have no further considerations to make. If, however, you could use either independent or correlated groups, you must make that decision before proceeding.

Control Issues As with the two-group designs discussed in Chapter 10, your decision to use the multiple-independent-groups design versus the multiple-correlated-groups design revolves around control issues. The multiple-independent-groups design uses the control technique of randomly assigning participants to groups. If you have a substantial number of research participants (at least 10 per group), you can be fairly confident that random assignment will create equal groups.

Multiple-correlated-groups designs use the control techniques of matching, repeated measures, or natural pairs to assure equality of groups and to reduce error variability. Recall the equation that represents the general formula for a statistical test:

$$\text{statistic} = \frac{\text{between-groups variability}}{\text{error variability}}$$

Reducing the error variability in the denominator of the equation will result in a larger computed statistical value, thereby making it easier to reject the null hypothesis. We hope you remember from Chapter 10 that using a correlated-groups design reduces your degrees of freedom, which makes it somewhat more difficult to achieve statistical significance and reject the null hypothesis. The reduced error variability, however, typically more than offsets the loss of degrees of freedom. Correlated designs therefore often produce stronger tests for finding statistical significance.

Practical Considerations Matters of practicality become quite important when we contemplate using a multiple-correlated-groups design. Let's think about each type of correlated design in turn. If we intend to use *matched sets*, we must consider the potential difficulty of finding three (or more) participants to match on the extraneous variable we choose. Suppose

we conduct a learning experiment and thus wish to match our participants on IQ. How difficult will it be to find three, four, five, or more participants (depending on the number of levels we use) with the same IQ? If we cannot find enough to make a complete set of matches, then we cannot use those participants in our experiment. We may, therefore, lose potential research participants through the requirement of large matched sets. We may be limited in our use of *natural sets* by set size also. How much chance would you have of running an experiment on triplets, quadruplets, or quintuplets? For this reason, using animal littermates is probably the most common use of natural sets in multiple-group designs. When we use *repeated measures* in a multiple-group design, we are requiring each participant to be measured at least three times. This requirement necessitates more time for each participant or multiple trips to the laboratory, conditions the participants may not be willing to meet. We hope this message is clear: If you intend to use a multiple-correlated-groups design, plan it very carefully so that these basic practical considerations do not sabotage your experiment.

What about practical considerations in multiple-independent-groups designs? The multiple-independent-groups design is simpler than the correlated version. The practical factor you must take into account is the large number of research participants you will have to make random assignment feasible *and* to fill the multiple groups. If participants are not available in large numbers, you should consider using a correlated design.

Drawing a definite conclusion about running independent- versus correlated-multiple-group designs is not simple. The correlated designs have some statistical advantages, but they also require you to take into account several practical matters that may make using such a design difficult. Independent designs are simple to implement, but they force you to recruit or obtain many research participants to assure equality of your groups. The best advice we can provide is to remind you that each experiment presents you with unique problems, opportunities, and questions. You should be aware of the factors we have presented and to weigh them carefully in conjunction with your experimental question when you choose a specific research design for your experiment.

Variations on the Multiple-Group Design

In Chapter 10 we discussed two variations on the two-group design. Those same two variations are also possible with the multiple-group design.

Comparing Different Amounts of an IV This “variation” on the multiple-group design is not actually a variation at all; it is part of the basic design. Because the smallest possible multiple-group design would consist of three treatment groups, every multiple-group design must compare different amounts (or types) of an IV. Even if a multiple-group design has a control group, there are at least two different treatment groups in addition.

If we already know that a particular IV has an effect, then we can use a multiple-group design to help us define the limits of that effect. In this type of experiment we often add an important control in order to account for a possible **placebo effect**.

For example, is it possible that some of the effects of coffee on our alertness are due to what we *expect* the coffee to do? If so, a proper control group would consist of people who drink decaffeinated coffee. These participants would be blind to the fact that their coffee does not contain caffeine. This group, without any caffeine, would show us whether coffee has any placebo effects.

Placebo effect An experimental effect caused by expectation or suggestion rather than the IV.

Dealing With Measured IVs All the research examples we have cited in this chapter deal with manipulated IVs. It is also possible to use measured IVs in a multiple-group design. In

Chapter 4 you learned that the research we conduct with a measured rather than a manipulated IV is termed **ex post facto research**. Remember that we cannot randomly assign participants to groups in such research because they already belong to specific groups. Thus, the groups may be different at the beginning of the research. We cannot draw cause-and-effect relations from such an experiment because we do not directly control and manipulate the IV ourselves. Still, an ex post facto design can yield interesting information and, because the design does use some controls, we may be able to rule out some alternative explanations. Let's look at a student example of an ex post facto design with a measured IV.

Radha Dunham and her advisor Lonnie Yandell (2005), of Belmont University in Nashville, Tennessee, used an ex post facto approach in their study of students' feeling of self-efficacy about their drawing ability. They chose their participants from an advanced art class, a lower level art class, and a general psychology class to represent advanced, beginning, and non-art groups, respectively. Why does this experimental design fit the multiple-group format? Does it have one IV? Yes: level of art skill. Does it have three or more levels of that one IV? Yes: the advanced, beginning, and non-art groups. These levels were their measured IV—the researchers could not assign students to one of the art skill groups; they could only "measure" which class students were taking as their approximation of level of art skill.

Dunham and Yandell (2005) found that the non-art group showed the lowest level of art self-efficacy compared to both the advanced and beginning art groups. The beginning and advanced art groups did not differ in their art self-efficacy. Notice that the multiple-group design allowed Dunham and Yandell to detect a difference between one group versus each of the two other groups. This type of findings shows the advantage of the multiple-group design over the two-group design; two two-group experiments would have been necessary to obtain the results from this one multiple-group experiment. Remember these various types of differences because we will return to them in the "Statistical Analysis" section of the chapter, next.

■ REVIEW SUMMARY

1. Psychologists plan their experiments beforehand using an **experimental design**, which serves as a blueprint for the experiment.
2. You use the **multiple-group design** for experimental situations in which you have one **independent variable** that has three or more levels or conditions.
3. A multiple-group design may or may not use a **control group**. If there is a control group, there are at least two experimental groups in addition.
4. You form **independent groups** of research participants by **randomly assigning** them to treatment groups.
5. You form **correlated groups** of research participants by creating matched sets, using natural sets, or measuring each participant more than once (repeated measures).
6. **Multiple-correlated-groups designs** provide extra advantages for experimental control relative to **multiple-independent-groups designs**.

7. Practical considerations in dealing with research participants make the multiple-correlated-groups designs considerably more complicated than multiple-independent-groups designs.
8. Multiple-group designs exist primarily to allow comparisons of different amounts (or types) of IVs.
9. Measured IVs can be used in multiple-group designs, resulting in **ex post facto studies**.

Check Your Progress

1. Why is the two-group design the building block for the multiple-group design?
2. The simplest possible multiple-group design would have _____ IV(s) and _____ treatment group(s).
3. What advantage(s) can you see in using a multiple-group design rather than a two-group design?
4. Devise an experimental question that could be answered with a multiple-group design that you could not answer with a two-group design.
5. Why are matched sets, repeated measures, and natural sets all considered *correlated-groups* designs?
6. What is the real limit on the number of groups that can be included in a multiple-group design? What is the practical limit?
7. Make a list of the factors you would consider in choosing between a multiple-group design and a two-group design.
8. Correlated-groups designs are often advantageous to use because they _____.
9. Why are practical considerations of using a multiple-correlated-groups design more demanding than those when using a two-correlated-groups design or a multiple-independent-groups design?
10. If we wished to compare personality traits of firstborn, lastborn, and only children, what type of design would we use? Would this represent a true experiment or an ex post facto study? Why?

Statistical Analysis: What Do Your Data Show?

We will remind you from the previous chapter that experimental design and statistical analysis are intimately linked. You *must* go through the decision-making process we have outlined before you begin your experiment in order to avoid the possibility that you will run your experiment and collect your data only to find out that there is no statistical test that you can use to analyze your data.

Analyzing Multiple-Group Designs

In this chapter we have looked at designs that have one IV with three (or more) groups. In your introductory statistics course you probably learned that researchers analyze these

One-way ANOVA A statistical test used to analyze data from an experimental design with one independent variable that has three or more groups (levels).

Completely randomized ANOVA This one-way ANOVA uses independent groups of participants.

Repeated-measures ANOVA This one-way ANOVA uses correlated groups of participants.

multiple-group designs with the *analysis of variance* (ANOVA) statistical procedure. As you will see, we will also use ANOVA to analyze designs that include more than one IV (see Chapter 12); hence, there are different types of ANOVAs, and we need some way to distinguish among them. In this chapter we are looking at an ANOVA for one IV; researchers typically refer to this procedure as a **one-way ANOVA**.

You remember that we have considered both multiple-independent-groups and multiple-correlated-groups designs in this chapter. We need two different types of one-way ANOVA to analyze these two types of designs, just as we needed different *t* tests in Chapter 10. As you can see from Figure 11-1, when we assign our participants to multiple groups randomly, we will analyze our data with a one-way ANOVA for independent groups (also known as a **completely randomized ANOVA**). On the other hand, if we use matched sets, natural sets, or repeated measures, we will use a one-way ANOVA for correlated groups (also known as a **repeated-measures ANOVA**) to evaluate our data.

Planning Your Experiment

In Chapter 10 we featured the statistical analysis of data from an experiment designed to compare the response time of salesclerks as a function of their customers' clothing (also see Chapter 9). That example, of course, cannot serve as a data analysis example for this chapter because it represents a two-group design.



Suppose we have already conducted the sample experiment covered in Chapters 9 and 10. How could we conduct a similar experiment using a multiple-group design?

Operational definition
Defining the independent, dependent, and extraneous variables in terms of the operations needed to produce them.

Between-groups variability Variability in DV scores that is due to the effects of the IV.

The *most* similar experiment would be one in which students in the introductory class dressed in three different types of clothing rather than just two. Suppose that we decide to investigate further because we found (in Chapter 10) that salesclerks responded more quickly to customers in dressy clothes than to those dressed in sloppy clothes. We decide to add an intermediate clothing group—we choose to add casual clothing as our third group. Again, we must consider the **operational definition** of our new IV group. We define casual clothing as slacks and shirts (e.g., khakis and polo shirts) for both male and female customers. We have 24 students in the class, so we have 8 students as "stimuli" in each of the three groups (one group for each type of clothing). We have the students go to the same store on the same day and randomly choose a department in which to browse. The store is large and employs many clerks, so there is no problem finding a different clerk for each student. This random choice will allow the salesclerks to be randomly assigned to the three groups (a requirement to create independent groups). An observer goes with the students to time unobtrusively the salesclerks' response time to each student, which is the dependent variable (DV). You can see the clerks' response times in Table 11-1. Let's discuss the basis behind the ANOVA procedure before we look at our statistical analyses.

TABLE 11-1

Salesclerks' Response Times (in Seconds) for Hypothetical Clothing Style Experiment

Clothing Styles		
Dressy	Sloppy	Casual
37	50	39
38	46	38
44	62	47
47	52	44
49	74	50
49	69	48
54	77	70
69	76	55
Mean = 48.38		Mean = 63.25
		Mean = 48.88

Rationale of ANOVA

We expect that you learned something about ANOVA in your statistics course. We introduced a closely related concept in the Control Issues section in Chapter 10; you may wish to refer back to that section. You will remember that variability in your data can be divided into two sources: **between-groups variability** and **error variability** (also known as **within-groups variability**). The between-groups variability represents the variation in the DV that is due to the IV; the error variability is due to such factors as individual differences, errors in measurement, and extraneous variation. In other words, error variability refers to *any* variability in the data that is not a product of the IV. Look at Table 11-2, which is a slightly altered version of Table 11-1.

Error variability

Variability in DV scores that is due to factors other than the IV, such as individual differences, measurement error, and extraneous variation (also known as *within-groups variability*).

Within-groups variability

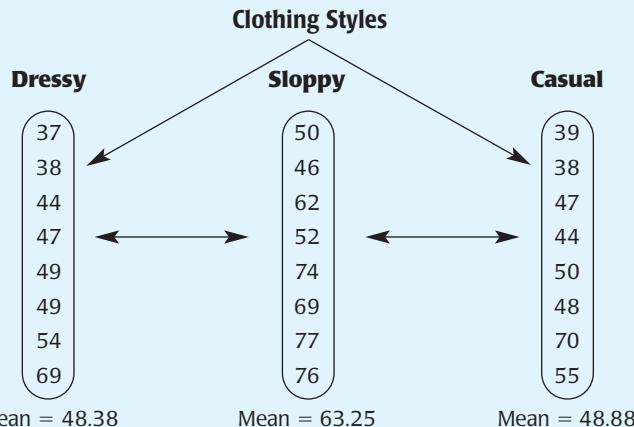
Another term for error variability.



What type of variability do you think is represented by the arrows in Table 11-2 (see next page)? What type of variability do you think is shown in the circled columns? If you have trouble with these questions, reread the previous paragraph.

The variability *between* the groups' response times represents the variability caused by the IV (the different types of clothing); therefore, the arrows represent the between-groups variability. If the times differ among the three groups of clerks, the clothes should be responsible for the difference (assuming we have controlled extraneous variables). On the other hand, error variability should occur among the participants *within* each particular group (thus its name, *within-groups variability*); this is the variability represented by the circled columns. One major source of within-groups variability is the differences within all the participants in a group—what we have labeled individual differences. Different people (or nonhuman animals) will score differently on the DV simply because they are different organisms.

TABLE 11-2 Salesclerks' Response Times (in Seconds) for Hypothetical Clothing Style Experiment



"Wait a minute," you may say. "What we have just described as within-groups variability—individual differences, measurement errors, extraneous variation—can occur between the groups just as easily as within the groups." This thought represents very good thinking on your part. Your point is correct and is well taken. Thus, we must change the formula that we reviewed just a few pages ago:

$$\text{statistic} = \frac{\text{between-groups variability}}{\text{error variability}}$$

The fact that we can find error between our groups as well as within our groups forces us to alter this formula to the general formula shown below for ANOVA. The *F* symbol is used for ANOVA in honor of Sir Ronald A. Fisher (1890–1962), who developed the ANOVA (Spatz, 2001).

$$F = \frac{\text{variability due to IV + error variability}}{\text{error variability}}$$

If our IV has a strong treatment effect and creates much more variability than all the error variability, we should find that the numerator of this equation is considerably larger than the denominator (see Figure 11-4A). The result, then, would be a large *F* ratio. If, on the other hand, the IV has absolutely no effect, there would be no variability due to the IV, meaning we would add 0 for that factor in the equation. In such a case, our *F* ratio should be close to 1 because the error variability between groups should approximately equal the error variability within groups. This situation is depicted in Figure 11-4B.

The notion that has evolved for the ANOVA is that we are comparing the ratio of between-groups variability (variability caused by the IV) to within-groups variability. Thus, the *F* ratio is conceptualized (and computed) with the following formula:

$$F = \frac{\text{between-groups variability}}{\text{within-groups variability}}$$

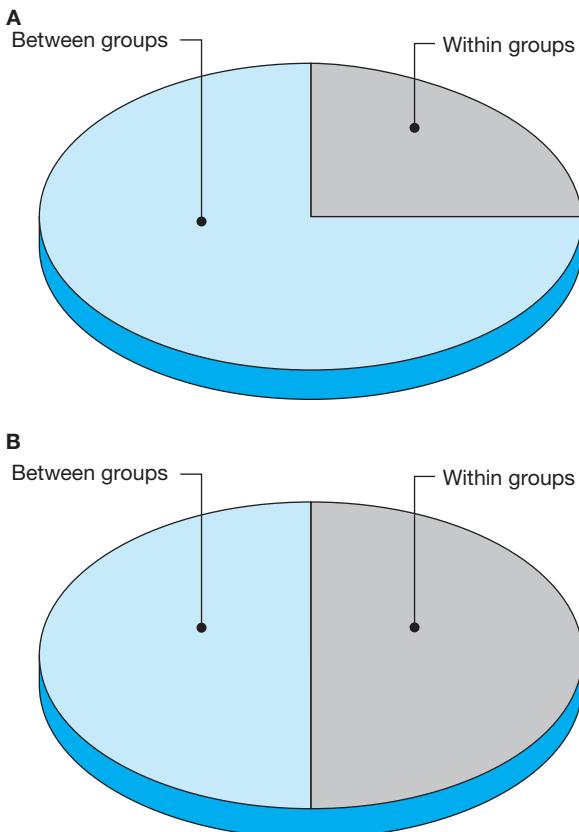


FIGURE 11-4 Possible Distributions of Variability in an Experiment. **A** depicts a large F ratio; **B** depicts an F ratio of 1.

A simple way to think of ANOVA is to realize that you are dividing the treatment effect by the error. When the IV has a significant effect on the DV, the F ratio will be large; when the IV has no effect or only a small effect, the F ratio will be small (near 1). You may wish to place a bookmark at this page—we will refer back to it shortly.

Interpretation: Making Sense of Your Statistics

With the addition of a third group, our experimental design has become slightly more complicated than the two-group design discussed in Chapter 10. As you will see, adding a third group (or more) creates an interesting statistical problem for us; we may have to compute an extra statistical test to explore significant findings. (Just in case you ever see such an analysis, one-way ANOVA can also be used for experiments with only two groups. We presented only the t test for such designs in the previous chapter to minimize overlap and possible confusion.)

Interpreting Computer Statistical Output

Once again we will look at generic computer output to give you experience with typical output so that you can better generalize your knowledge to the particular statistical package that is available to you. The results appear in Table 11-3.

One-Way ANOVA for Independent Samples We are examining results from a one-way ANOVA because we have one IV with three groups. We used the ANOVA for independent samples because we randomly assigned salesclerks to the three different clothing conditions. The DV scores represent the clerks' response times to differently attired customers.

As usual, we first look for information about descriptive statistics. You will find the descriptive statistics in the top portion of Table 11-3. Before going on, remember that we recommended that you make sure you have entered your data correctly in the computer by checking the means using a calculator. It will take only a few minutes but will spare you from using an incorrect set of results if you somehow goofed when you put the numbers in the computer. We can see that Group 1 (clerks responding to customers in dressy clothes) had a mean response time of 48.38 seconds, Group 2 (sloppy clothing) had a mean of 63.25 seconds, and Group 3 (casual clothes) responded in 48.88 seconds on the average. So, we do see numerical differences among these means, but we do not know whether the differences

TABLE 11-3 Computer Output for One-Way ANOVA for Independent Samples

GROUP	N	MEAN	STD DEV	STD ERR	95% CONF INT
1 (dressy)	8	48.38	10.11	3.57	39.92-56.83
2 (sloppy)	8	63.25	12.54	4.44	52.76-73.74
3 (casual)	8	48.88	10.20	3.61	40.34-57.41
ONEWAY ANOVA: RESPTIME by CLOTHING					
SOURCE	SUM OF SQUARES	DF	MEAN SQUARES	F RATIO	PROB.
BETWEEN GROUPS	1141.75	2	570.88	4.71	.02
WITHIN GROUPS	2546.25	21	121.25		
TOTAL	3688.00	23			
POST HOC TEST: Tukey-HSD with significance level .05					
* Indicates significant differences shown for pairings					
	G r p 1	G r p 2	G r p 3		
Mean	CLOTHING				
48.38	Grp 1				
63.25	Grp 2	*		*	
48.88	Grp 3				

are large enough to be significant until we examine the inferential statistics. We also see the standard deviation and standard error (standard deviation divided by \sqrt{n}) for each group (the times for group 2 are more variable than those of the other two groups), as well as 95% confidence intervals. You may remember that confidence intervals provide a range of scores between which μ (the true population mean) should fall. Thus, we are 95% confident that the interval of 40.34 to 57.41 seconds contains the population mean for all clerks responding to customers in casual clothing.

The inferential statistics appear immediately below the descriptive statistics. We see the heading “ONEWAY ANOVA,” which lets us know that we have actually computed a one-way ANOVA. The subheading shows us that we have analyzed the variable “RESPTIME” in relation to the “CLOTHING” variable. This label simply means that we have analyzed our DV (RESPTIME, the clerks’ response times) by our IV (CLOTHING, the three styles of dress).

The output from ANOVA is typically referred to as a **source table**. In looking at the table, you will see “SOURCE” printed on the left side of the page. Source tables get their name because they isolate and highlight the different *sources* of variation in the data. In the one-way ANOVA table, you see two sources of variation: between groups and within groups.

Source table A table that contains the results of ANOVA. *Source* refers to the source of the different types of variation.



The two terms “between groups” and “within groups” refer to what?

Between groups is synonymous with our treatment (IV) effect, and *within groups* is our error variance. The **sum of squares**, the sum of the squared deviations around the mean, is used to represent the variability of the DV in the experiment (Kirk, 1968). We use ANOVA to divide (partition) the variability into its respective components, in this case between-groups and within-groups variability. In Table 11-3 you see that the total sum of squares (variability in the entire experiment) is 3688, which we partitioned into between-groups sum of squares (1141.75) and within-groups sum of squares (2546.25). The between-groups sum of squares added to the within-group sum of squares should always be equal to the total sum of squares ($1141.75 + 2546.25 = 3688$).

If we formed a ratio of the between-groups variability and the within-groups variability based on the sums of squares, we would obtain a ratio of less than 1. We cannot use the sums of squares for this ratio, however, because each sum of squares is based on a different number of deviations from the mean (Keppel, Saufley, & Tokunaga, 1992). Think about this idea for a moment: Only three groups can contribute to the between-groups variability, but many different participants can contribute to the within-groups variability. Thus, to put them on an equal footing, we have to transform our sums of squares to **mean squares**. We make this transformation by dividing each

Sum of squares The amount of variability in the DV attributable to each source.

Mean square The “averaged” variability for each source; computed by dividing each source’s sum of squares by its degrees of freedom.

sum of squares by its respective degrees of freedom. Because we have three groups, our between-groups degrees of freedom are 2 (number of groups minus 1). Because we have 24 participants, our within-groups degrees of freedom are 21 (number of participants minus the number of groups). Our total degrees of freedom are equal to the total number of participants minus 1, or 23 in this case. As with the sums of squares, the between-groups degrees of freedom added to the within-groups degrees of freedom must equal the total degrees of freedom ($2 + 21 = 23$). Again, our mean squares are equal to each sum of squares divided by its degrees of freedom. Thus, our between-groups mean square is 570.88 ($1141.75/2$), and our within-groups mean square is 121.25 ($2546.25/21$).

Variance A single number that represents the total amount of variation in a distribution; also the square of the standard deviation, σ^2 .

We should note at this point that a mean square is analogous to an estimate of the **variance**, which you may remember from statistics as the square of the standard deviation (σ^2). As soon as we have the mean squares, we can create our distribution of variation. Rather than drawing pie charts, like those shown in Figure 11-4, we compute an *F* ratio to compare the two sources of variation. Referring to the bookmark we advised you to use a few pages back, we find that the *F* ratio is equal to the between-groups variability divided by the within-groups variability. Because we are using mean squares as our estimates of variability, the equation for our *F* ratio becomes

$$F = \frac{\text{mean square between groups}}{\text{mean square within groups}}$$

Thus, our *F* ratio of 4.71, as shown in Table 11-3, was derived by dividing 570.9 by 121.3. This result means that the variability between our groups is almost five times larger than the variability within the groups. Or, perhaps more clearly, the variability caused by the IV is almost five times larger than the variability resulting from error. If we drew a pie chart for these results, it would look like Figure 11-5.

Finally, we come to the conclusion (or so we think!). Did the different clothing styles have a significant effect? Next to "F RATIO" in Table 11-3 you see the "PROB" entry: .02. This probability of chance of these data (if the null hypothesis is true) is certainly lower than .05, so we did find a significant difference. The difference in the response times among the three groups of salesclerks probably did not occur by chance. Although the computer printed the probability of

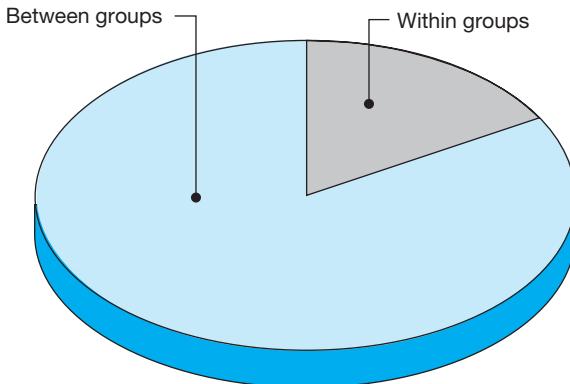


FIGURE 11-5 Distribution of Variability for Different Clothing Experiment.

chance for you, you should know how to use a printed F table just in case your computer program does not calculate probabilities. This activity is somewhat different from using a printed t table because in ANOVA we have two values for degrees of freedom. In this case our degrees of freedom are 2 (between groups, in the numerator) and 21 (within groups, in the denominator). When we look in the F table (see Table A-2), the column shows the numerator df and the row shows the denominator df . In this case you must find 2 in the numerator column and 21 on the denominator line and locate the intersection of those two numbers. At that point you will see 3.47 as the .05 cutoff and 5.78 as the .01 cutoff. Because our F value of 4.71 is between these two numbers, the probability that it could have occurred by chance is less than .05 but greater than .01. Thus, if you were using a table rather than the computer output, you would have written $p < .05$ or $.01 < p < .05$ (to be as specific as possible). Sherlock Holmes stated, "I could only say what was the balance of probability" (Doyle, 1927, p. 93).

With the two-group design, we would be finished with our computer output at this point and could go on to interpreting our statistics in words. With significant findings in a two-group design, we merely note the higher mean and conclude that it is significantly higher than the other mean. This decision procedure is not correct, however, with the multiple-group design when we find a significant difference because we have more than two means. We know there is significance among our means because of our significant F ratio, but which one(s) is (are) different from which one(s)? From a significant F ratio, we cannot tell.

To discern where the significance lies in a multiple-group experiment, we must conduct additional statistical tests known as **post hoc comparisons** (also known as *follow-up tests*). These tests allow us to determine which groups differ significantly from each other after we have determined that there is overall significance (by finding a significant F ratio). Many different post hoc tests exist, and there is much debate over these tests that is beyond the scope of this text. Simply remember that you will have to conduct post hoc tests if you find overall significance in a one-way ANOVA.

Post hoc comparisons

Statistical comparisons made between group means after finding a significant F ratio.

At the bottom of Table 11-3, you see the results of a post hoc test known as the Tukey HSD (an abbreviation for *honestly significant difference*). The Tukey test allows you to test all pair-wise comparisons, meaning that you can test the difference between all sets of two means (Keppel et al., 1992). In looking at Table 11-3, we see that Group 2 is significantly different at the .05 level from both Group 1 and Group 3 according to the Tukey test. This result means that the clerks took significantly longer to wait on sloppily dressed students (63.25 seconds) than either those in casual (48.88 seconds) or those in dressy clothes (48.38 seconds). No other groups differed significantly from each other, meaning that there was no statistical difference in the time it took clerks to wait on students dressed well or casually.

As in the previous chapter, we hope that you are learning general principles about computer printouts rather than specific words or terms for which you will blindly search. If you understand the general principles, interchanging *between groups* with *clothing* (or the name of some other IV) should not be problematical for you; different statistical programs may simply use different ways of getting at the same thing (much like having slightly different names for the same test). For example, don't be surprised to see the label *error* rather than *within groups*—both terms mean the same thing. The important conclusion is that given the same data, any two programs should find the same results.

Translating Statistics Into Words Let us remind you, as we did in Chapter 10, that the results of any statistical test are only as good as your experimental procedures. In other words, if you have conducted a sloppy experiment, your statistical results will be meaningless. When we draw the conclusion that our IV has caused a difference in the DV scores, we are assuming that we conducted a well-controlled experiment and removed extraneous variables from the scene. If you find that extraneous variables have confounded your experiment, you should not interpret your statistics because they are now meaningless. For the same reason, detectives must learn specific ways to collect evidence in the field. If they collect contaminated evidence, all the lab tests in the world cannot yield a definitive conclusion.

Based on our inferential statistics, we can conclude that the clothing customers wear is important because clerks took differing amounts of time to wait on customers depending on how they were dressed.



Although this conclusion is technically correct, it is a poor conclusion. Why? How would you revise this conclusion to make it better?

This conclusion is poor because it is incomplete. Reread the sentence and decide what you can learn from it. All you know is that students who wore some type of clothing were waited on more quickly than other students who wore some other clothing. Thus, you know that clothes can make a difference, but you don't know which type of clothing led to faster responses. To write a good conclusion, we must go back to our inferential statistics, particularly the post hoc tests. In Table 11-3 we find that the students wearing dressy clothes were waited on in 48.38 seconds, students wearing sloppy clothes received attention in 63.25 seconds, and students wearing casual clothes got help in 48.88 seconds. The significant F ratio lets us know that there is significance *somewhere* among those means. The Tukey post hoc comparison tests informed us that the differences between Group 2 and both Groups 1 and 3 were significant. To interpret this difference, we must examine our descriptive statistics. When we examine the means, we are able to conclude that students in shabby clothes got help significantly more slowly than the students in dressy or casual clothes. *No other mean differences were significant.*

We must determine how to communicate our statistical findings in APA format. We will use a combination of words and numbers. There are many different ways to write this set of findings in an experimental report. Here is one example:

The effect of different clothing on salesclerks' response time was significant, $F(2, 21) = 4.71$, $p = .02$. The proportion of variance accounted for by the clothing effect (η^2) was .31. Tukey tests indicated ($p < .05$) that clerks waiting on customers dressed in sloppy clothes ($M = 63.25$, $SD = 11.73$) responded more slowly than clerks waiting on customers in dressy ($M = 48.38$, $SD = 9.46$) or casual clothes ($M = 48.88$, $SD = 9.55$). The response times of clerks waiting on customers in dressy and casual clothes did not differ from each other.

The words alone should convey the meaning of our results. Could someone with no statistical background read and understand these sentences if we removed the numbers? We think so. The inferential test results explain our findings to readers with a statistical background. The descriptive statistics allow the reader to observe exactly how the groups performed and how variable that performance was. The effect size information reported here, η^2 (eta squared), is similar to r^2 because it tells you the proportion of variance in the DV (response times) accounted for by

the IV (clothing). (An easy way to calculate η^2 is to divide the between-groups sum of squares by the total sum of squares.) The reader has an expectation about what information will be given because we write our results in this standard APA format. You will find this type of communication in results sections in experimental articles. As you read more results sections, this type of communication will become familiar to you.

One-Way ANOVA for Correlated Samples Now we will look at the one-way ANOVA for correlated samples. The sample experiment about clothing and clerks' response times we have used so far in the chapter fits the multiple-group design for independent samples and thus is not appropriate to analyze with the one-way ANOVA for correlated samples.



How could you modify the experiment concerning salesclerks' reactions to customers' style of dress so that it used correlated groups rather than independent groups?

To be correct, you should have proposed the use of matched sets, natural sets, or repeated measures in your modified experiment. The best choices would involve matched sets or repeated measures; we don't think that natural sets is a feasible choice in this situation—you're not using littermates, and finding sets of triplets who are all salesclerks is most unlikely! If you choose matched sets, you must decide on a matching variable. It is difficult to know what variable on which you should match salesclerks that would be related to their reactions to different clothing styles. Matched sets would not be a good choice for forming correlated groups.

Imagine that you conduct your experiment at a small store that employs only a few salesclerks. In order to have enough data points, you decide to have each salesclerk respond to a customer in each of the three clothing groups. Because you would measure each clerk's response times to all three styles of dress, you would control possible individual differences between the clerks. (Another scenario that might lead you to use repeated measures would occur if you decided that it is likely that some variable in salesclerks, such as attitude, might affect their response times to customers in different types of clothing. Using repeated measures would allow you essentially to cancel out differences between different clerks, because each clerk would wait on a customer in each clothing group.)

You are now ready to begin the experiment. The students each dress in one of the three styles and enter the store. Because we are using repeated measures, we know for certain that the clerks in the three groups are equated (because they are the same clerks in each group). Given this hypothetical example, the scores in Table 11-1 would now represent sets of response times from eight salesclerks. (Remember that in the real world, it is not legitimate to analyze the same data with two different statistical tests. This is a textbook, certainly not the real world, and we are doing this as an example only.)

You can see the results for the one-way ANOVA for correlated samples in Table 11-4. As usual, we are first interested in examining descriptive statistics. The descriptive output is shown at the top of Table 11-4. As you can see, we obtain the mean, standard deviation, sample size, and 95% confidence interval for each group. The descriptive statistics for the three groups match what we have previously seen in Table 11-3, which is certainly logical. Although we are now using a correlated-samples analysis, nothing has changed about the samples themselves. So, we see the same means, standard deviations, and confidence intervals that we have seen before.

TABLE 11-4 Computer Output for One-Way ANOVA for Correlated Samples

GROUP	N	MEAN	STD DEV	STD ERR	95% CONF INT
1 (dressy)	8	48.38	10.11	3.57	39.92-56.83
2 (loose)	8	63.25	12.54	4.44	52.76-73.74
3 (casual)	8	48.88	10.20	3.61	40.34-57.41
ONEWAY ANOVA: RESPTIME by CLOTHING (CORR SAMP)					
SOURCE	SUM OF SQUARES	DF	MEAN SQUARES	F RATIO	PROB.
CLOTHING	1141.75	2	570.88	19.71	.000
SUBJECTS	2140.65	7	305.81	10.56	.000
WITHIN CELLS	405.59	14	28.97		
TOTAL	3688.00	23			
POST HOC TEST: Tukey-HSD with significance level .01					
* Indicates significant differences shown for pairings					
	G r p 1	G r p 2	G r p 3		
Mean	CLOTHING				
48.38	Grp 1				
63.25	Grp 2				
48.88	Grp 3				

The other information that we see in Table 11-4 is our ANOVA source table. Once again, the entries in this particular source table vary slightly from the tables we've looked at earlier. Although you may begin to believe that this is some sinister plot hatched just to confuse you, you have to focus on the basic information, remembering that terms are used slightly differently in different situations. Here, our three sources of variance are labeled "within cells," "subjects," and "clothing." Because you know that we are comparing different types of clothing as our IV, it should be clear that "clothing" represents the effect of our IV and "within cells" represents our source of error variation (refer back to Table 11-2 to see our within-cell variation pictorially represented by the circles). "Subjects," of course, represents the variability between different salesclerks. When we examine the source table, we find that the *F* ratio for the comparison of the clerks' response times to different clothing styles is 19.71, with 2 (numerator) and 14 (denominator) degrees of freedom, which results in a probability of chance of .000 according to the computer.

This situation illustrates one of our pet peeves with computerized statistical programs.

When you studied distributions in your statistics class, what did you learn about the tails of those distributions? We hope you learned that the tails of distributions are **asymptotic**; that is, the tails extend into infinity and never touch the baseline. This fact means that the probability of a statistic is *never* .000. No matter how large the statistic gets, there is always some small amount of probability under the tails of the distribution. Unfortunately, people who design statistics software either

Asymptotic Refers to tails of distributions that approach the baseline but never touch the baseline.

have a limited number of columns to work with or they don't think about this issue, so they have the computer print a probability of .000, implying that there is no uncertainty. In light of this problem, we advise you to list $p < .001$ if you ever find such a result on your computer printout.

Pardon the digression, but you know how pet peeves are! Back to the statistics. The overall effect of the clothing is significant, which leads us to wonder which clothing styles differ from each other. This source table looks different from the one in Table 11-3 because it shows the effects of two IVs: CLOTHING and SUBJECTS. Although the SUBJECTS effect is significant, it does not tell us anything very important: We simply learn that there were significant differences between the eight salesclerks' response times. In other words, we found individual differences between the salesclerks. This effect is an expected one and is not profound. Typically, you would ignore this effect. However, the SUBJECTS effect is important statistically. If you compare Tables 11-3 and 11-4, you will see that the correlated samples ANOVA has taken the SUBJECTS variability (mean square) out of the WITHIN CELLS (or error) variability compared to the WITHIN GROUPS (or error) term in the independent-samples ANOVA. This difference demonstrates the power of the correlated-samples analysis to reduce variability in the error term and to create a larger F ratio.

As with the multiple-group design for independent samples, we used a Tukey test for post hoc comparisons. Again, we found that Group 2 (sloppy clothing) was significantly different ($p < .01$) from both Group 1 (dress clothes) and Group 3 (casual clothes); however, Groups 1 and 3 did not perform significantly differently from each other. Notice that our significant differences are at the .01 level rather than .05 as with the independent-samples case. This change is another indication of the increased power of the correlated-samples analysis.

Translating Statistics Into Words Our experimental logic is no different from that for the independent-samples ANOVA. The only difference is that we used a somewhat more stringent control procedure in this design—we used repeated measures with our participants rather than assigning them to groups randomly.

Our conclusions should combine our numbers with words to give the reader a clear indication of what we found. Remember to include information both about any difference that was found and the directionality of the difference.



How would you write the results of this experiment in words and numbers for an experimental report?

Although the conclusion for the correlated-samples test is similar to that for the independent-groups test, it is different in some important ways. We hope you figured out those important differences. Here's a sample conclusion:

The effect of three different clothing styles on clerks' response times was significant, $F(2, 14) = 19.71, p < .001$. The proportion of variance accounted for by the clothing effect (η^2) was .74. Tukey tests showed ($p < .01$) that clerks took longer to respond to customers dressed in sloppy clothes ($M = 63.25, SD = 11.73$) than to either customers in dressy clothes ($M = 48.38, SD = 9.46$) or customers in casual clothes ($M = 48.88, SD = 9.55$). Response times did not differ between the clerks waiting on customers in dressy or casual clothing.

Did your conclusion look something like this? Remember, the exact wording may not necessarily match—the important thing is that you cover all the critical details.



There are five important differences between this conclusion and the conclusion drawn for the independent-groups ANOVA. Can you find them?

The *first* difference comes in the degrees of freedom. There are fewer degrees of freedom for the error term in the correlated-groups ANOVA (**WITHIN CELLS**) than for the independent-samples case (**WITHIN GROUPS**). *Second*, the *F* value for the correlated-groups test is larger than for the independent-groups test. The larger *F* value is a result of reducing the variability in the denominator of the *F* equation. This difference in *F* values leads to the *third* difference, which is the probability of chance. Despite the fact that there are fewer degrees of freedom for the correlated-samples case, its probability of chance is lower (smaller). *Fourth*, the proportion of variance accounted for by the clothing effect (η^2) was considerably larger. *Fifth*, the post hoc tests show a lower (smaller) probability of chance in the correlated-groups situation.

These last three differences most clearly show the advantage of a correlated-groups design. Because using repeated measures reduced some of the error variability, the probability of the difference coming about by chance is smaller than it was in the independent-samples case. Thus, the conclusion from the correlated-groups design yields the clearer finding (reducing the chance of a Type I error). We cannot promise that correlated-groups designs will always allow you to find a clearer difference than independent-groups designs; however, we can tell you that correlated-groups designs do increase your odds of detecting smaller significant differences because such designs reduce error variance.

The Continuing Research Problem

In Chapters 9 and 10 we began our continuing research project by looking at clerks' response times as a function of how customers were dressed. Clerks' times were significantly higher when they waited on customers in sloppy clothing than when they waited on well-dressed customers. Because of this result, we decided to pursue this line of research further and, in this chapter, compared the effects of three different styles of clothing to each other. On the basis of our results, we can state that salespeople wait on customers in dressy or casual clothing more quickly than they wait on customers in sloppy clothing.

Is our research project complete at this point? As you might have realized, we could compare an endless number of styles of dress. This research problem could go on forever. In all seriousness, you might have wondered about the effects of other possible IVs on salesclerks' response times. As we begin to ask more complicated questions, we must move on to more complex designs to handle those questions. In Chapter 12 we will be able to continue our research problem with an experiment using more than one IV at a time.

Let's review the logical steps we took in conducting this experiment. Refer back to Figure 11-1 to take note of our experimental design questions.

1. After conducting a preliminary experiment (Chapter 10) and determining that salesclerks waited on well-dressed customers more quickly, we decided to test further the effects of different clothing (IV) on clerks' response times (DV).

2. We chose to test only one IV (clothing) because our research is still preliminary.
3. We tested three different styles of dress because they seemed to be valid ways for customers to dress.
- 4a. With access to many salesclerks, we used random assignment to the three groups and, thus, a multiple-independent-groups design. We used a one-way ANOVA for independent groups and found that clerks responded more quickly to customers in dressy or casual clothes than to customers in sloppy clothes.
- 4b. With smaller numbers of clerks, we chose to use repeated measures. Thus, we used a multiple-within-group design and a one-way ANOVA for correlated groups. Clerks responded to sloppily dressed customers more slowly than to well-dressed or casually dressed customers.
5. We concluded (hypothetically) that customers should not dress in sloppy clothes if they desire to get helped quickly in a store.

■ REVIEW SUMMARY

1. When your experimental design consists of one IV with three or more groups and you have randomly assigned participants to groups, the proper statistical analysis is a one-way ANOVA for independent groups (**completely randomized ANOVA**).
2. When your experimental design has one IV with more than two groups and you have used matched sets, natural sets, or repeated measures, you should analyze your data with a one-way ANOVA for correlated groups (**repeated-measures ANOVA**).
3. ANOVA partitions the variability in your DV into **between-groups variability** (caused by the IV) and **within-groups variability** (resulting from sources of error). We then compute a ratio between these two sources of variation known as the *F* ratio.
4. ANOVA results are typically shown in a **source table**, which lists each source of variance and displays the *F* ratio for the effect of the IV.
5. A significant *F* ratio merely indicates that there is a significant difference somewhere among your various groups. **Post hoc comparisons** are necessary to determine which groups differ from each other.
6. Using APA format for our statistical results allows us to convey our findings in both words and numbers in a clear and concise manner.
7. Previous experiments often lead to further questions and new experiments. The multiple-group design is an ideal design to follow up on the results from a two-group experiment.

■ Check Your Progress

1. Suppose you wish to compare the ACT or SAT scores of the freshman, sophomore, junior, and senior classes at your school to determine whether differences exist among those students. Draw a block diagram of this design. What design and statistical test would you use to conduct this research?

2. You wonder whether students who take the ACT or SAT three times are able to improve their scores significantly. You select a sample of such students and obtain their three scores. What type of experimental design does this question represent? Draw a block diagram of it. What statistical test would you use to analyze the data?
3. When we look at our F ratio and its probability in a multiple-group design, why can't we examine the descriptive statistics directly to reach a conclusion about our experiment?
4. The variability that is due to our IV is termed the _____ variance, whereas the variability caused by individual differences and error is the _____ variance.
5. Suppose you conducted the experiment summarized in Question 2 and found the following statistics: $F(2, 24) = 4.07, p < .05$. On the basis of this information, what could you conclude?
6. What additional information do you need in Question 5 to draw a full and complete conclusion?
7. In the continuing research problem from this chapter, why was it important to have the (hypothetical) knowledge from the similar study in Chapter 10?
8. You decide to test how people's moods vary by the four seasons. What type of experimental design would you use for this research project? Why?
9. You choose to test people's preferences for fast-food hamburgers, and you have McDonald's, Burger King, Wendy's, and White Castle franchises in your town. What type of experimental design would you use for this research project? Why?

■ Key Terms

Experimental design, 231	Matching variable, 236	Within-groups variability, 245
Independent variable, 232	Placebo effect, 241	Source table, 249
Principle of parsimony, 233	Ex post facto research, 242	Sum of squares, 249
Levels, 233	One-way ANOVA, 244	Mean square, 249
Treatment groups, 233	Completely randomized ANOVA, 244	Variance, 250
Independent groups, 235	Repeated-measures ANOVA, 244	Post hoc comparisons, 251
Correlated groups, 235	Operational definition, 244	Asymptotic, 254
Random assignment, 235	Between-groups variability, 245	
Control procedure, 235	Error variability, 245	
Confounded experiment, 235		

■ Looking Ahead

In this chapter we furthered our knowledge about research design and how it fits with particular experimental questions. Specifically, we looked at an extension of the basic building-block design by using one IV and three or more groups. In the next chapter we will make a significant alteration in our basic design by adding a second IV. This expanded design will give us the ability to ask much more sophisticated questions about behavior because most behaviors are affected by more than one variable at a time.

Designing, Conducting, Analyzing, and Interpreting Experiments with Multiple Independent Variables

Experimental Design: Doubling the Basic Building Block

- The Factorial Design
- Comparing the Factorial Design to Two-Group and Multiple-Group Designs
- Choosing a Factorial Design
- Variations on Factorial Designs

Statistical Analysis: What Do Your Data Show?

- Naming Factorial Designs

- Planning the Statistical Analysis
- Rationale of Factorial ANOVA
- Understanding Interactions

Interpretation: Making Sense of Your Statistics

- Interpreting Computer Statistical Output
- A Final Note

The Continuing Research Problem

Experimental Design: Doubling the Basic Building Block

This chapter will continue building on the experimental design material that we first encountered in Chapter 10. We will see many familiar concepts from the two previous chapters, as well as from earlier in the text (e.g., independent, dependent, and extraneous variables, control, and so on). You can expect that some of this chapter will be a review of familiar concepts; however, we will apply those concepts in a new and different framework—the factorial experimental design.

Let's return to our analogy for experimental design that we first saw in Chapter 10: building objects with Legos or Tinkertoys. Chapters 10 and 11 presented the beginner's set and a slightly larger version of that beginner's set, respectively. Now, with the factorial design, we encounter the top-of-the-line, advanced set that has all the possible options. When you buy the largest set of building materials, you can build anything from very simple objects to extremely complicated structures; the same is true of experimental design. Using a factorial design gives us the power we need to investigate several **factors** or **independent variables (IVs)** in a single experiment. Factorial designs are the lifeblood of experimental psychology because they allow us to look at *combinations* of IVs at the same time, a situation that is quite similar to the real world. A factorial design is more like the real world because there are probably few, if any, situations in which

Factors Synonymous with IVs.

Independent variables
Stimuli or aspects of the environment that are directly manipulated by the experimenter to determine their influences on behavior.

your behavior is affected by only a single factor at a time. Imagine trying to isolate only one factor that affected your ACT or SAT scores! What about intelligence, motivation, courses you had taken, a test time of 8:00 on Saturday morning, health, and so on?

In Chapter 11 we used another analogy for our experimental designs; we compared the two-group design to a house, the multiple-group design to an apartment building, and the factorial design to a skyscraper. The idea behind this analogy was not to frighten you into worrying about how complex factorial designs are but instead to make two points. First, as we have already mentioned, even complex designs are based on the principles that you previously encountered with simple designs. As you move from building a house to a skyscraper, most of the principles of building remain the same—they are merely used on a larger scale. This formula is true also for experimental design: Although we will be dealing with more complex designs in this chapter, you already have the majority of the background you need from Chapters 10 and 11. Second, just as building larger buildings gives you more decisions and options, designing larger experiments gives you more decisions and options. Decisions, of course, imply responsibility. You will have to take on additional responsibility with a factorial design. Rather than planning an experiment with only one IV, you will be planning for two or three (or possibly even more) IVs. Additional IVs mean that you have more factors to choose and control. By taking on additional responsibilities, you also gain additional information. By moving from a one-IV experiment to a two-IV experiment, you will gain information about a second IV *and* the interaction between the two IVs. We will soon discuss interactions in depth. Let's examine the factorial design first.

The Factorial Design

In this chapter we will expand the basic two-group design by doubling it. Look back at Figure 10-2 (page 207) and imagine what would happen if you doubled it. Of course, one possible result of doubling Figure 10.2 would be a design that we have already covered in Chapter 11: a multiple-group design with four levels. That design would result in an experiment similar to the continuing research problem of Chapter 11 if we planned an experiment contrasting the effects of four different styles of clothing rather than three. This type of doubling, however, would be an extension rather than an expansion. We would take one IV with two levels (dressy clothes and shabby clothes) in Chapter 10 and extend it from two to four levels.



What would an *expansion* (rather than an extension) of Figure 10-2 look like? You may wish to draw a sketch of your answer.

Does your drawing resemble Figure 12-1? Contrast Figure 12-1 with Figures 11-2 and 11-3 (pages 234 and 235). Do you see the difference? To what can you attribute this difference? You'll find the answer in the next section.

How Many IVs? In the preceding two chapters the first question we faced when considering the choice of an experimental design dealt with how many IVs our experiment would have. Nothing has changed in Chapter 12—we still begin with “How many independent

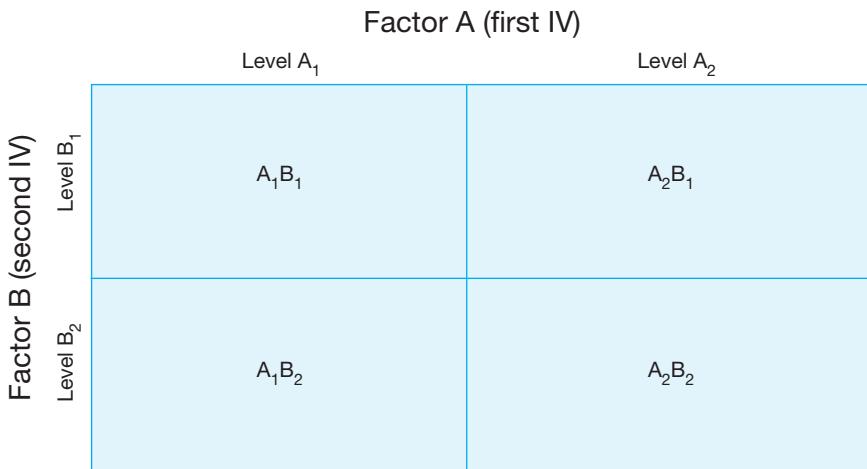


FIGURE 12-1 Simplest Possible Factorial Design (2×2).

variables will I use in my experiment?" (see Figure 12-2). For the first time, though, we have a different answer. We are moving into more complex designs that use at least two IVs. We refer to any design that uses at least two IVs as a **factorial design**. The factorial design gets its name because we refer to each IV as a factor; multiple IVs, then, yield a factorial design. Theoretically, there is no limit to the number of IVs that you can use in an experiment. Practically speaking, however, it is unlikely that you would want to design an experiment with more than two or three IVs. After that point the increased complexity is such that it would tax your ability to conduct the experiment and your skill to interpret the results. Hence, we will use experimental designs with two IVs to illustrate our points (we give you an illustration of a design with three IVs in Appendix C). Real-life detectives often face complex cases that are similar to factorial designs. Imagine that you are a detective confronted with several suspects and that you are faced with evaluating the means, motive, and opportunity for each suspect. Can you see how such a situation is analogous to a factorial design?

Factorial design An experimental design with more than one IV.

How Many Groups or Levels? In looking at Figure 12-2, you will notice that this question that we asked in both Chapters 10 and 11 does not appear on the relevant portion of the flowchart. The reason for its absence is simple: Once you have two (or more) IVs, you will use a factorial design. The number of levels for each factor is unimportant at this point.

Let's return to Figure 12-1 for a moment. Do you see how it represents two two-group designs stacked on top of each other? That is what we meant earlier by doubling or expanding the two-group design. If we take two two-group designs and combine them, we end up with the design pictured in Figure 12-1, which has two IVs, each with two levels. This figure represents the simplest possible factorial design, which is known as a 2×2 *design*. This 2×2 shorthand notation tells us that we are dealing with a design that has two factors (IVs) because there are two digits given and that each of the two factors has two levels because each digit shown is a 2. In other words, the *number* of numbers tells us how many IVs there are; the *value* of each number tells us how many levels each IV has. Finally, when we complete

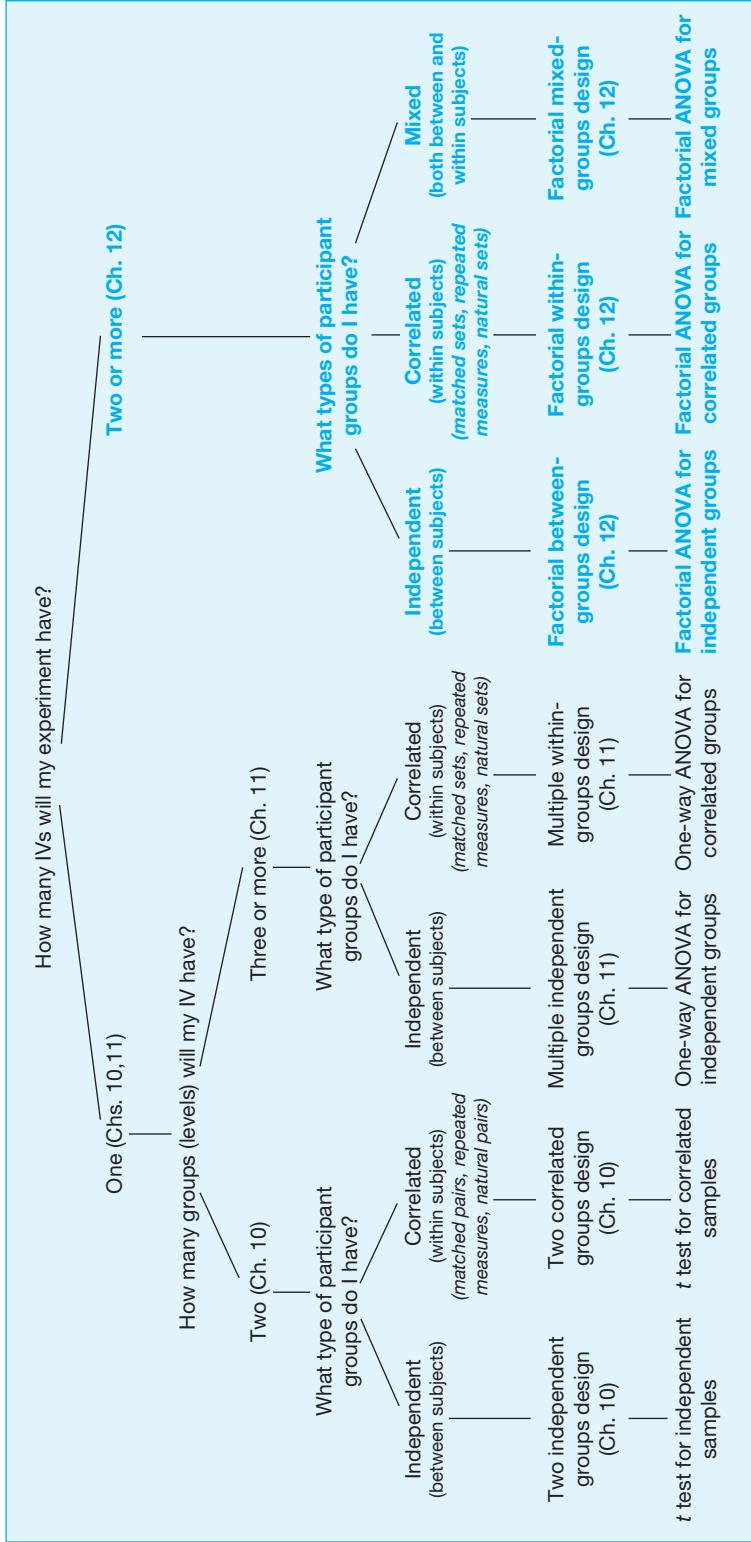


FIGURE 12-2 Experimental Design Questions.

the implied multiplication, this notation also tells us how many unique treatment combinations (or groups) our experiment requires. A 2×2 design requires four treatment combinations (2 times 2), whereas a 3×3 design would require nine treatment combinations, and a 2×4 design would have eight treatment combinations.



Suppose you designed an experiment that had one IV with two levels, one IV with three levels, and another IV with four levels. What would be the shorthand notation for this design?

Because there are three IVs, there will be three digits in the answer. Each digit represents the number of levels of one of the IVs. Thus, your answer should have been "a $2 \times 3 \times 4$ design."

Figure 12-1 shows an additional design notation. Various factors are often designated by letters, so the first factor is labeled Factor A, the second Factor B, and so on. The levels within a factor are often designated by the letter that corresponds to the factor *and* with a number to differentiate the different levels. Thus, the two levels within the first factor would be labeled A_1 and A_2 . If the factor has more than two levels, we continue numbering them in similar fashion until we reach the last level (e.g., $A_1, A_2, A_3, \dots, A_n$), where n represents the number of levels of the factor.

What might such a 2×2 design look like in real life? To begin answering this question, refer to Figure 10-2 (page 207), where we diagrammed the two-group experiment of McClellan and Woods (2001), in which they compared the response times of salesclerks to deaf and hearing customers.



How could you expand McClellan and Woods's (2001) experiment so that it becomes a 2×2 factorial? Draw a diagram.

First, we hope you drew a design like the one shown in Figure 12-3. The first IV (A) should be customer hearing, with the two levels representing the experimental group (customers who were deaf) and the control group (customers who were not deaf). The second IV (B) could be any possible variable, as long as the two groups differ on that IV. For example, let's say that you are interested in determining whether the effects of the customers' hearing status vary based on the salesclerks' sex (as we mentioned in Chapter 10). Your second IV (a measured rather than manipulated IV) would be the sex of the clerks, with your two groups composed of women versus men. This particular experimental design appears in Figure 12-3. Four treatment groups are necessary, one for each possible combination of the two levels of the two treatments. Thus, we would have a group of female salesclerks who wait on deaf customers, a group of female salesclerks who wait on hearing customers, a group of male salesclerks who wait on deaf customers, and a group of male salesclerks who wait on hearing customers.

		Factor A (customer hearing)	
		Deaf Customers	Hearing Customers
Factor B (salesclerk sex)	Women	deaf customers waited on by female clerks	hearing customers waited on by female clerks
	Men	deaf customers waited on by male clerks	hearing customers waited on by male clerks

FIGURE 12-3 Expansion of McClellan and Woods's (2001) Hearing Disability Study.

We hope it is clear to you that this 2×2 design is composed of two two-group designs. One two-group design contrasts the effects of customers' hearing, and the second two-group design contrasts the effects of salesclerk sex. At this point, you might ask, "Wouldn't it be simpler to run two separate experiments rather than combining them both into one experiment?" Although it might be somewhat easier to run two experiments, there are two disadvantages of conducting separate experiments. First, it is not as time efficient as running one experiment. Even if you used the same number of participants in the experiment(s), two experiments will simply take longer to complete than one; there are fewer details to deal with in one experiment than in two. Imagine how inefficient it would be for a detective to complete an investigation of one suspect before even beginning the investigation of a second suspect! Second, by running two experiments, you would lose the advantage that you gain by conducting a factorial experiment: the interaction.

Main effect Refers to the sole effect of one IV in a factorial design.

Interaction The joint, simultaneous effect on the DV of more than one IV.

When we combine two IVs in one experiment, we get all the information we would get from two experiments, plus a bonus. In the example building on McClellan and Woods's (2001) experiment, we will still determine the solitary effect of customer hearing *and* the solitary effect of salesclerk sex. In a factorial experiment, these outcomes from the IVs are termed the **main effects**. The bonus that we get from a factorial experiment is the **interaction** of the two IVs. We will discuss interactions at length later in this chapter, but let us provide you a preview at this point. Suppose we ran the customer hearing/salesclerk sex study and found the results shown in Figure 12-4.



Can you interpret the main effects in Figure 12-4? Did the customers' ability to hear have any effect? Did the sex of the salesclerk have any effect? Study that graph carefully to answer these questions.

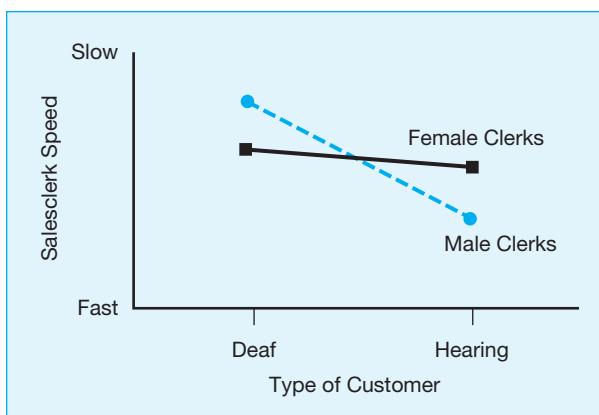


FIGURE 12-4 Hypothetical Results of Customer Hearing/Salesclerk Sex Experiment.

It does appear that the customer hearing factor had an effect on clerks' response times. Both groups of clerks (men and women) seemed to respond more slowly when the customer was deaf; notice how the lines for both groups go down as they move from left (deaf customers) to right (hearing customers). On the other hand, the sex of the salesclerks seems to have made no *overall* difference in their time to help. It appears that both men and women waited on customers in approximately the same amount of time. If you average the two points on the women's line and average the two points on the men's line, the two means are virtually identical.

The crossing pattern you see graphically portrayed in Figure 12-4 illustrates an interaction effect. We find significant interactions when the effects of one IV change as the level(s) of the other IV changes. Another common way of describing interactions is that *the effects of one IV depend on the particular level of another IV*. If these descriptions are difficult for you to understand, look at Figure 12-4 and ask yourself whether the effects of the customer's hearing are the same for both groups of salesclerks. Although the response times of both sexes increased when they waited on customers who were deaf, it is clear that this effect is much more pronounced for the male clerks, who were much slower in waiting on deaf customers than hearing customers. On the other hand, the female clerks responded similarly to both deaf and hearing customers, although they were slightly slower in waiting on customers who were deaf. This difference for female clerks might not even be significant. If you were describing these results to someone, would it be correct to say merely that salesclerks were slower when they waited on deaf customers? Or should you note that the effect of customer hearing seemed to vary as a function of the sex of the clerks—that the customer's hearing greatly affected the response time in men but had only a small or nonexistent effect for women? It is apparent that the second conclusion, although more complex, paints a much clearer, more accurate picture of the results. Thus, it seems that the effects of customer hearing depend on the sex of the salesclerks; the effect of one IV depends on the specific level of the other IV. Let's practice some more with interpreting factorial results to make sure that you understand interactions.



Can you interpret the main effects *and* interactions in Figures 12-5 and 12-6? Did the customer's ability to hear have any effect? Did the sex of the salesclerk have any effect? Was there an interaction? Study the graphs carefully to answer these questions.

Let's look at Figure 12-5 first. To answer the question about customer hearing, you have to figure out what the mean for the deaf and hearing customers would be. Notice that if you average the two points for each group, the overall means would be nearly equal; the salesclerks waited on both deaf and hearing customers in a moderate amount of time. In similar fashion, to determine the effect of salesclerk sex, we have to find the means for the female and male clerks. Again, the average response time for the clerks does not seem to differ by gender; female and male clerks took about the same amount of time to wait on customers. The simplest way to discern an interaction is to look for lines on the graph that are *not* parallel. The lines in Figure 12-5 are clearly not parallel. Thus, the effects of one IV depend on the level of the second IV. In this hypothetical case, female clerks waited on deaf customers more quickly than hearing customers, but male clerks waited on hearing customers more quickly than deaf customers. Note that you have to include both IVs in your explanation of the significant interaction. Figure 12-5 shows both main effects to be nonsignificant and the interaction to be significant. How did you do at deciphering those results?

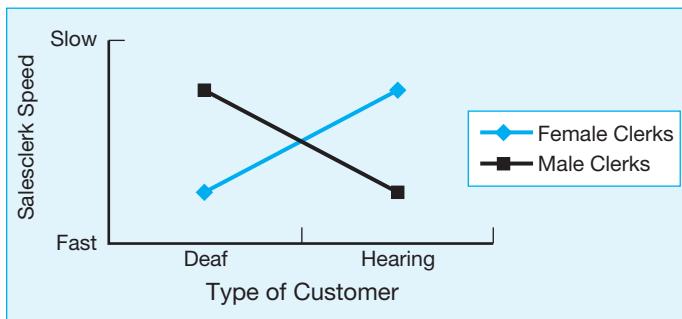


FIGURE 12-5 Hypothetical Results of Customer Hearing/Salesclerk Sex Experiment.

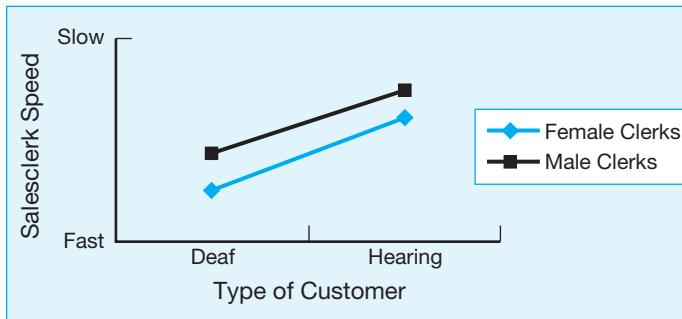


FIGURE 12-6 Hypothetical Results of Customer Hearing/Salesclerk Sex Experiment.

Now we turn to Figure 12-6; we will go through the same set of questions. If we examine the means based on the type of customer, we find that deaf customers received help more quickly than hearing customers. Also, we see that female clerks tended to give help more quickly than male clerks. Thus, both main effects show significance. How about the interaction? The lines look roughly parallel, don't they? Notice that we can explain each main effect without having to refer to the second IV: Deaf customers received help more quickly (regardless of sex of the salesclerk). Female clerks were faster than male clerks (regardless of the customers' hearing). Therefore, the interaction is not significant in Figure 12-6.

We have not covered all the possible outcomes of a two-variable experiment. With two main effects and an interaction, any possible combination of significant and nonsignificant effects can occur. All three can be significant, all three can be nonsignificant, or any combination can occur. It is critical that you can look at such graphs and interpret them. We will return to interactions later in this chapter to make certain that you understand them fully.

Let's look at an actual student example of a factorial design. S. Merle Riepe (2004), a student at Nebraska Wesleyan University in Lincoln, Nebraska, was interested in studying the effects of education level and gender on achievement motivation. Riepe used participants from three groups for differing education levels: freshmen and sophomores, juniors and seniors, and managers from a hospital. In addition, each group was made up of both men and women for the gender variable (of course). Riepe asked each participant to complete a questionnaire that measured the need for both achievement and affiliation; Riepe did not calculate the affiliation scores. This combination of treatments appears in Figure 12-7. As you can see, there are two IVs (education level and gender), with the education level having three levels and gender two levels.

Participants began the experiment by completing a demographic form. Participants then completed the questionnaire to measure achievement motivation. Riepe's results appear in Figure 12-8. The graph shows that the participants' level of education had a significant effect on their achievement motivation; the upper level students had lower motivation scores than the managers. The lower level students were intermediate and did not differ from either of

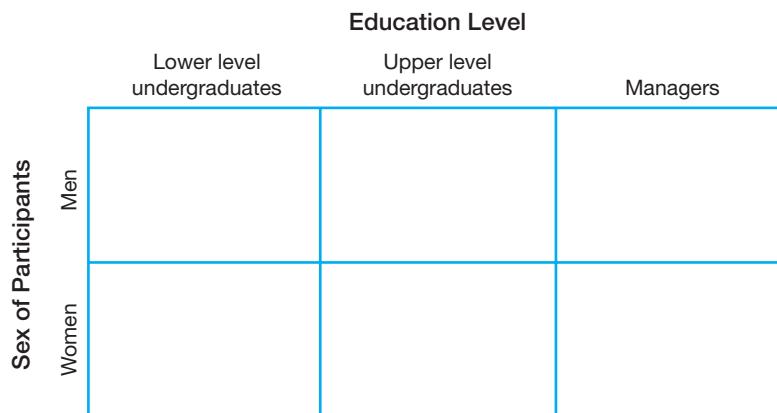


FIGURE 12-7 Diagram of Riepe's (2004) Experiment.

Source: Adapted from Figure 1 from "Effects of Education Level and Gender on Achievement Motivation," by S. M. Riepe, 2004, *Psi Chi Journal of Undergraduate Research* 9, pp. 33–38. Copyright © 2004 Psi Chi, The National Honor Society in Psychology (www.psichi.org). Reprinted by permission. All rights reserved.

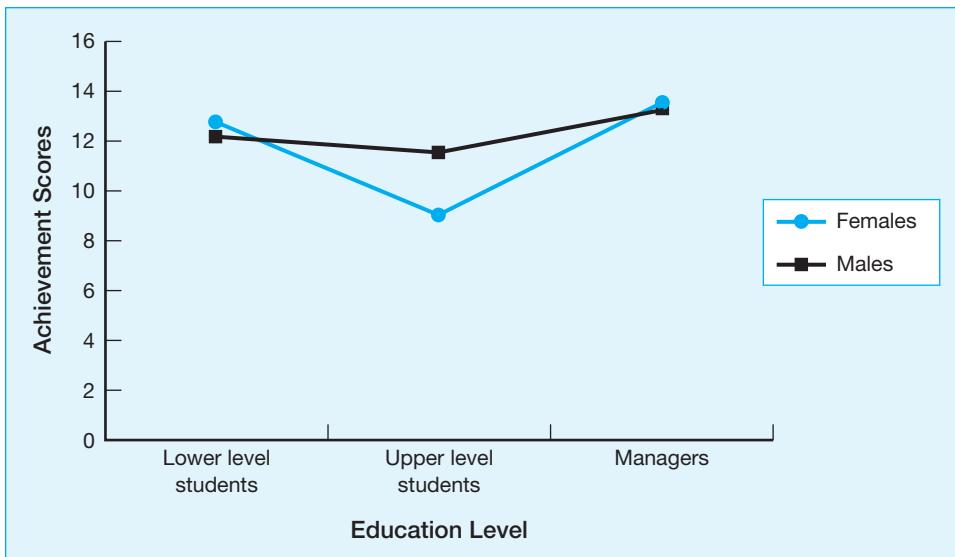


FIGURE 12-8 Results from Riepe's (2004) Experiment.

Source: Adapted from Figure 1 from "Effects of Education Level and Gender on Achievement Motivation," by S. M. Riepe, 2004, *Psi Chi Journal of Undergraduate Research* 9, pp. 33–38. Copyright © 2004 Psi Chi, The National Honor Society in Psychology (www.psichi.org). Reprinted by permission. All rights reserved.

Independent groups

Groups of participants formed by random assignment.

Correlated groups

Groups of participants formed by matching, natural pairs, or repeated measures.

Mixed assignment A factorial design that has a mixture of independent groups for one IV and correlated groups for another IV. In larger factorial designs at least one IV has independent groups and at least one has correlated groups (also known as *mixed groups*).

the other two groups. There was no difference in the achievement motivation scores between men and women or between participant gender and education level (the interaction). Although the two lines on the graph are not perfectly parallel, they do not deviate significantly from that orientation. We can conclude, therefore, that achievement motivation varied only as a function of education level, regardless of the sex of the participant.

Assigning Participants to Groups As in Chapters 10 and 11, the matter of how we assign our research participants to groups is important. Again, we have two options for this assignment: **independent groups** and **correlated groups**. This question is not answered this time as simply as it was in the two-group and multiple-group designs, each of which had only one IV. Matters can become more complicated in a factorial design because we have two (or more) IVs to consider. All IVs could have participants assigned randomly or in a correlated fashion, or we could have one IV with independent groups and one IV with correlated groups. We refer to this last possibility as a **mixed assignment** (or mixed groups).

Random Assignment to Groups Factorial designs in which *both* IVs involve random assignment may be called *between-subjects factorial designs* or *completely randomized designs* (see Figure 12-2). These labels should make sense from our discussion of between- and within-groups variance in Chapter 11. Between-groups variance refers to variability between independent groups of participants, which results from random assignment.

Riepe (2004) used a completely randomized design for his experiment dealing with achievement motivation as a function of education level and sex. He randomly

assigned participants to one of the six possible treatment groups based on education level (lower level undergraduate/upper level undergraduate/manager) and sex (male/female). Thus, his experiment required the use of six independent groups of participants; he used six groups ranging from 13 to 31 participants. Each group was independent of the others in the experiment.

After he collected the data, Riepe compared the achievement scores for participants *between* those who had varying degrees of education and *between* participants of differing sexes. The comparisons were between-groups comparisons because Riepe had different participants with no relationships in each group.

Nonrandom Assignment to Groups In this section we examine factorial designs in which participant groups for all IVs have been formed through nonrandom assignment. We refer to such designs as *completely within-groups* (or *within-subjects*) designs. We may want to resort to nonrandom assignment in order to assure the equality of participant groups before we conduct the experiment. It is important to remember, particularly in research with small numbers of participants or with small IV effects, that random assignment cannot guarantee equality of groups. Let's take a brief look at the three methods of creating correlated groups.

1. Matched Pairs or Sets. Matching can take place in either pairs or sets because factorial designs can use IVs with two or more levels. The more levels an IV has, the more work matching for that variable takes, simply because of having to form larger sets of participants that are equated on the matching variable. Also, the more precise the match that is necessary, the more difficult the matching becomes. For example, matching on sex or educational level may be fairly simple; however, matching on college major or family background may be more difficult. Likewise, when more than one IV uses matching, the demand for a large number of a specific type of research participant may become overwhelming. For instance, using matched participants for both IVs in a 2×2 experiment requires four matched groups, a 2×3 requires six such groups, and a 3×3 design would necessitate nine matched groups. Imagine what would happen in an experiment with three IVs! For this reason, factorial designs that use matching for all IVs are relatively rare.

2. Repeated Measures. You will remember that a repeated-measures design is one in which participants take part in all levels of an IV. In a completely within-groups experiment using repeated measures, participants would take part fully and completely; that is, they would participate in every possible treatment combination. As you might imagine, this requirement often makes it difficult or impossible to conduct an experiment with repeated measures on multiple IVs.



Would it have been possible to conduct Riepe's (2004) experiment with repeated measures on both IVs? If it were possible, would it be wise to do so? Why or why not?

To answer the first question, you have to understand its implication. Look at Figure 12-7. To conduct this experiment using repeated measures on both IVs, each participant would have to experience each of the six possible conditions. Could a single participant participate in each of the educational level conditions? Certainly yes, but only in a longitudinal study

(see Chapter 4). Given that Riepe was an undergraduate who wanted to finish his project in a single semester or year, the answer to the previous question is a resounding no. Could a single participant participate in both participant sex conditions? Only in a very unusual situation! Thus, it really would not be possible for Riepe to have repeated measures on both IVs.

Even if it were *possible* to use repeated measures on both variables, would it be wise to do so? “Being wise” is a little harder question to answer than “being possible” because it involves a value judgment. If you can envision a problem with a design, it probably is *not* wise to use that design. For example, it would often be logical to assume that participating in all conditions of an experiment would alert participants to the two (or more) independent variables in the experiment. If the participants can figure out what an experiment is all about, then you have to worry about their expectancies and demand characteristics (see Chapter 7), which can affect their responses in the experiment. In some cases, therefore, even if it is possible to use repeated measures for the independent variables, it may not be wise to do so.

Of course, not all factorial experiments have such problems with using repeated measures. It is possible to design an experiment with multiple IVs in which you expose participants to all treatment combinations. As you have probably guessed by now, the smaller the design, the more feasible it is to include the same participants in all conditions of the experiment. Thus, a 2×2 design would be the most likely candidate for a totally repeated-measures design.

3. Natural Pairs or Sets. Using natural groups in a totally within-subjects design has the same difficulties as the matched pairs or sets variation of this design, but it would be even harder. If it would be difficult to form matched pairs or sets to participate in a factorial design, imagine how much harder it would be to find an adequate number of naturally linked participants. At least when we match participants, we usually have a large number to measure in our attempt to make matches. To use natural sets, we would have to find substantial numbers of natural groups. As we noted in Chapter 11, experimenters rarely use this approach other than with animal littermates. We will not consider natural sets in this chapter because of their infrequent use.

Let’s examine a student research project that used a factorial design with a totally within-subjects approach. Gale Lucas, Bennet Rainville, Priya Bhan, Jenna Rosenberg, and Kari Proud, students at Willamette University in Salem, Oregon, and their faculty advisor, Susan Koger, were interested in memory distortions for visual stimuli. In particular, Lucas et al. (2005) were interested in knowing whether boundary extension (the tendency for viewers to “remember” an image with larger boundaries than they originally saw) was affected by the type of image shown. In this study the researchers used photographs and computer-generated images (their first IV). In addition, participants saw a standard image, a zoomed-in version of the image, or a zoomed-out version of the image (their second IV). Thus, Lucas et al.’s design was a 2 (type of image) \times 3 (perspective of image) design.



Draw a block diagram of Lucas et al.’s experiment and label it completely.

		Perspective of Image		
		Standard	Zoomed In	Zoomed Out
Type of Image	Computer-Generated			
	Photograph			

FIGURE 12-9 Block Diagram of Lucas et al.’s (2005) Experiment.

Source: From “Memory for Computer-Generated Graphics: Boundary Extension in Photographic vs. Computer-Generated Images,” by G. M. Lucas, B. Rainville, P. Bhan, J. Rosenberg, K. Proud, & S. M. Koger, 2005, *Psi Chi Journal of Undergraduate Research*, 10, pp. 43–48.

We hope you drew a diagram like the one shown in Figure 12-9. If so, good job! If not, review the material earlier in this chapter concerning number of IVs and levels of IVs.

The participants in Lucas et al.’s (2005) experiment saw 12 standard images and then completed a survey about their visual preferences (as a distractor task). Following the survey, the participants saw replica images of the first 12 images, with a third being presented in the same manner, a third zoomed-in, and a third zoomed-out. Half of the images were photographs and half were computer-generated images. Participants had to indicate whether each of the replica images was the one that they had originally seen. The fact that participants saw images from all six possible treatment conditions means that Lucas et al. used repeated measures for both IVs; the participants viewed both types of images in all three perspectives. Lucas et al. found a significant interaction between type of image and image perspective such that participants were more likely to extend boundaries (report zoomed-out pictures as one of the originals) for computer-generated images than for actual photographs. Boundary restriction (reporting a zoomed-in picture as an original) did not differ between the two types of images. Had Lucas et al. conducted this experiment as two separate two-variable designs, they would never have gained the information about the interaction between type of image and perspective of image.

Mixed Assignment to Groups Because factorial designs have at least two IVs, we have the opportunity for a new type of group assignment that we did not encounter in the single-IV designs. As we mentioned previously, mixed-assignment designs involve a combination of random and nonrandom assignment, with at least one IV using each type of assignment to groups. In a two-IV factorial design, mixed assignment involves one IV with random assignment and one IV with nonrandom assignment. In such designs the use of repeated measures is probably more likely than other types of nonrandom assignment. What we often end up with in a *mixed factorial design* is two independent groups that are measured more than once (repeated measures). Such an experimental plan allows us to measure a difference between our groups and then determine whether that difference remains constant over time or across different types of stimuli. Mixed designs combine the advantages of the two types of designs. The conservation of participants through the use of repeated measures for a between-subjects variable makes for a popular and powerful design.

Let's consider a student research example that used a mixed design. Nida Ittayem, a student at Agnes Scott College in Decatur, Georgia, and Eileen Cooley, her faculty advisor, conducted a study to examine students' ability to lower their stress as a function of activities in which the students engaged. They had 34 female undergraduates participate in one of three conditions: A narrative writing group wrote about a significant emotional experience, a drawing group drew about a significant emotional experience, and a control group wrote about daily events without including emotions. Each participant engaged in her assigned activity for four consecutive days; Ittayem and Cooley (2004) asked each student to measure her level of stress each day after the writing or drawing exercise.



What are the two IVs in this experiment, and what are their respective levels? Which is the between-groups variable and which is the within-groups variable? How would you describe this experiment in short-hand notation? What is the DV in this experiment?

One IV is the type of activity, which had three levels (writing emotion, drawing emotion, writing control). The second IV was time of stress measurement (the four consecutive days). Type of activity was the between-groups variable because participants engaged in only one activity. The three groups were independent—there was no relationship or pairing among the participants who used the different activities. Time was the within-groups variable—students participated in all conditions of this variable because they took the stress scale each day immediately after writing or drawing. Because this experiment had two IVs, one (type of activity) with three levels and one (time) with four levels, it was a 3×4 design (traditionally, in a mixed design, the between-subjects variable is listed first and the within-subjects variable is listed second). Finally, the DV was the students' ratings of their stress. The experimenters compared students' stress ratings between the three types of activities and for the four different times of measurement.

In their experiment Ittayem and Cooley (2004) found no significant main effect for either activity type or time; however, the interaction between the two variables was significant. Students who wrote or drew about emotional experiences showed a steady decline in stress over the four days. The students in the control writing group, on the other hand, showed an increase in stress from Day 1 to 2, a decrease in stress from Day 2 to 3, and a large increase in stress from Day 3 to 4. It is the different pattern of stress in the three groups across the four days that caused the interaction to be significant.

■ REVIEW SUMMARY

1. The plan psychologists use to guide their experiments is known as the **experimental design**.
2. If your experiment consists of two (or more) IVs, you will use a **factorial design**. Each IV (individually) is considered a **main effect**.
3. The number of levels of each IV is not important in choosing your particular factorial design.
4. Combining two IVs in an experiment allows you to test for **interactions**; that is, situations in which the effect of one IV depends on the specific level of another IV.

5. When you randomly assign research participants to their groups, you have **independent groups**. If all the IVs in your factorial design use independent groups, you are employing a between-subjects factorial design.
6. When you use matched pairs or sets, repeated measures, or natural pairs or sets, you have **correlated groups**. If all the IVs in your experiment use correlated groups, you are using a within-subjects factorial design.
7. Mixed factorial designs result from using both independent groups and correlated groups in a factorial design (**mixed assignment**). At least one IV must use independent groups and at least one must use correlated groups.

Check Your Progress

1. How is the two-group design related to the factorial design? Draw a picture as part of your answer.
2. Why is there a practical limit to the number of IVs you could use in an experiment?
3. You have conducted a 2×2 experiment; what information will you obtain from the analysis?
 - a. the effects of IV A
 - b. the effects of IV B
 - c. the effects of $A \times B$
 - d. all of the above
4. Matching
 1. mixed factorial design
 2. totally between-groups design
 3. totally within-groups design
 - A. fraternity members versus nonmembers; men versus women
 - B. fraternity members matched for family income measured twice
 - C. fraternity members versus nonmembers measured twice
5. The simplest possible factorial design would have _____ IV(s) and _____ total treatment group(s).
6. Devise an original example of a factorial design that uses mixed assignment to groups.

Comparing the Factorial Design to Two-Group and Multiple-Group Designs

At this point you have discovered that there are three major types of experimental designs: two-group, multiple-group, and multiple-IV design. Each of these three types of experimental designs has at least two subtypes, based on participant assignment variables (independent versus correlated groups, plus mixed assignment for multiple-IV designs). As a beginning researcher, you may find this array of experimental designs somewhat overwhelming when you plan your first experiment. To help you plan your research, we will cover some of the basic issues faced by researchers when they plan their studies.

As we mentioned earlier, factorial designs are based on the basic building-block designs we encountered in the two previous chapters. We create factorial designs by combining two of our basic building-block designs into a single design. Again, detectives use basic investigative principles as their building blocks even in complex cases.

You may remember that we described two-group designs as being ideal for a preliminary investigation of a particular IV in a presence-absence format (refer back to Figure 10-2, page 207). In a similar fashion, 2×2 factorial designs may be used for preliminary investigations of two IVs. If you look at Figure 12-3, where we created a hypothetical expansion of McClellan and Woods's (2001) hearing study, you can see that we used that design to make a preliminary investigation of salesclerks' responses based on customer hearing (deaf versus hearing) and sex of the clerks (female versus male). When we completed this experiment, we would have information about whether customer hearing and clerks' sex have any effects on clerks' response times. Suppose we wished to go further: What if we wanted to delve deeper into the effects of customer disability or sex?

In Chapter 11 we found that we could use the multiple-group design to conduct more in-depth investigations of an IV that interests us. We took our basic two-group design and extended it to include more levels of our IV. We can make the same type of extension with factorial designs. Figure 12-10 shows an extension of Figure 12-3, which includes three levels of one IV, thus creating a 3×2 factorial design (of course, there is no way to add a third level to salesclerks' sex). Notice that Figure 12-10 is simply a three-level, multiple-group experiment and a two-group experiment combined into one design. From this hypothetical design, we can get much more specific information about the effects of customer disability because we used three different types of customers with two types of disabilities rather than just the presence versus absence of a disability, as shown in Figure 12-3. Just as with the multiple-group design, there is no limit to the number of levels for any IV in a factorial design. Also, the number of levels of the IVs can be unequal (as in this case) or equal. Thus, we could create 2×5 factorial designs, 3×3 factorial designs, 3×6 factorial designs, and so on.

To this point, our discussion of Figure 12-10 has not added anything that we couldn't obtain by conducting two separate experiments. Whether we conduct one factorial experiment or two single-IV experiments, we will uncover information about the effects of our main effects or IVs. However, as we have already seen, the true advantage of factorial designs is

		Customer Disability Status		
		Deaf Customers	Blind Customers	Hearing Customers
Salesclerk Sex	Women			
	Men			

FIGURE 12-10 Combination of Two Multiple-Group Designs.

that they measure interaction effects. We have mentioned several times that interaction effects enable us better to understand the complexities of the world in which we live.

Unfortunately, we frequently come across students who are not interested in learning about complex relations as characterized by interaction effects. Instead, they would prefer to conduct single-IV experiments because the results from such experiments are simpler to understand. We are not against simple research designs—they contribute a great deal to our ability to understand the world—but we are against the attitude that says, “I want to use simple research designs because they are easier.” If your sole reason for choosing a research design is that it is simpler, we believe that you are making a choice for the wrong reason. Remember that we have cautioned you to choose the simplest research design that will *adequately test your hypothesis*. It is possible that the simplest research design available will *not* adequately test your hypothesis. For example, if we already have a great deal of information about a particular IV, then a presence-versus-absence manipulation of that IV is probably too simple. By the same token, if we already know that changes in a particular DV are caused by complex factors, then a simple design may not advance our knowledge any further. Let us provide you with an example in the following paragraph to illustrate what we’re talking about.

Let’s suppose that a friend, Joe, wishes to conduct research on why college students make the grades that they do. In this research, grades will be the DV, and Joe wants to isolate the IV that causes those grades. One logical choice for a factor that causes grades is intelligence because it stands to reason that more intelligent students will make higher grades and less intelligent students will make lower grades. Because Joe knows that he cannot give intelligence tests to a large number of college students, he uses ACT or SAT scores as a very rough measure of intelligence. Joe chooses a group of students with high ACT or SAT scores and a group of students with low ACT or SAT scores and compares their respective grades. Sure enough, Joe finds that the students in the high group have higher GPAs than those in the low group. Joe is excited because his hypothesis has been confirmed. He writes a report about his results and goes on his merry way, telling everyone he meets that intelligence causes differences in college students’ grades.



Is there anything wrong with the scenario we have just sketched for you? Do you see any specific flaws in the research itself? Do you see any flaws in the reasoning or in the original design?

Is this research flawed? Not really—there are neither obvious violations of experimental guidelines nor obvious extraneous variables. A better answer is that the original reasoning is flawed, or at least that it is too simplistic.

Think about the original question for a moment. Do you believe that ACT or SAT scores can be used to explain *everything* there is to know about college students’ grades? If you are like most students, your answer will be a resounding no. We all know students who entered college with low test scores, perhaps even on probation, but who have gone on to make good grades. On the other hand, we know students who entered with academic scholarships and yet flunked out. Clearly, there must be more to grades than intelligence or whatever

entrance exams measure. What about factors such as motivation and study skills? What about living in a dorm versus an apartment? What about being married versus being single? What about belonging to a sorority or fraternity versus being an independent? All these factors (and many more) could contribute to some of the variability in GPA that we observe among college students. Thus, if we decide to “put all our eggs in one basket”—entrance exam scores—we may be simplifying the question too much.

The problem with asking simple questions is that we get simple answers because that is all that we *can* get from a simple question. Again, there is nothing inherently wrong or bad about asking a simple question and getting a simple answer—*unless* we conclude that this simple answer tells us everything we need to know about our subject matter. In such a case we would be guilty of what Sherlock Holmes criticized: “You see, but you do not observe” (Doyle, 1927, p. 162). Asking more complex questions may yield more complex answers, but those answers may give us a better idea of how the world actually works. Factorial designs give us the means to ask these more complex questions.

Choosing a Factorial Design

Three considerations are important when you choose a particular factorial design. At the heart of the choice are your experimental questions; factorial designs provide considerable flexibility in devising an experiment to answer your questions. Second, it will not surprise you that you should consider issues of control in your design choice because experimental design is primarily concerned with the notion of control. Third, due to the wide degree of experimental choices possible with factorial designs, considerations of a practical nature are also important.

Experimental Questions The number of questions we can ask in a factorial experiment increases dramatically. Being able to ask additional questions is a great opportunity, but it also puts a burden on us. When we ask additional questions, we must make certain that the questions coordinate with each other. Just as many people would not want to wear clothes with colors that clash, we do not want to ask questions that “clash.” By clashing questions, we refer to questions that do not make sense when put together. No doubt you have sat in class and heard a student ask a question that seemed to have no relation to what was being covered in class. Experimental questions that have no relevance to each other may seem to clash when combined in the same experiment. For example, suppose you heard of a proposed experiment to find the effects of self-esteem and eye color on test performance. Does that planned experiment jar you somewhat? We hope it does. Does it make sense to combine self-esteem and eye color in an experiment? Does it make sense to examine the effects of eye color on test performance? This factor sounds like an IV that was thrown in simply because it could be. Could eye color be a logical IV? Perhaps it could in another situation. For example, eye color might well influence people’s judgments of a target person’s attractiveness or even of his or her intelligence. It seems unlikely, though, that eye color might affect one’s test performance. We hope that “off-the-wall” combinations of IVs will be minimized by a review of the existing psychological literature. When you base your experimental questions on previous research and theory, the odds of using strange combinations of IVs are decreased.

Control Issues We hope that by now you are able to anticipate the topic of discussion when you see the Control Issues heading. A glance at Figure 12-2 will remind you that we do have to consider independent versus correlated groups in factorial designs. A complicating

factor for factorial designs is that we have to make this decision for each IV we include in an experiment.

Research psychologists typically assume that random assignment to groups will adequately equate the groups if you have approximately 10 participants per group. On the other hand, a correlated assignment scheme (matching or repeated measures) will provide you with greater assurance of the equality of the groups. We hope that by now you fully understand the reasoning behind these two approaches. If you need a review, look back to Chapters 10 and 11.

Practical Considerations As IVs multiply in experimental designs, some of the practical issues involved become more complex. Often when students find that they can ask more than one question in an experiment, they go wild, adding IVs left and right, throwing in everything but the proverbial kitchen sink. Although curiosity is a commendable virtue, it is necessary to keep your curiosity somewhat in check when designing an experiment.

Remember the **principle of parsimony**, which we encountered in Chapter 10: You are well advised to keep your experiment at the bare minimum necessary to answer the question(s) that most interest(s) you. We heard this principle cast in a slightly different light when a speaker was giving advice to graduate students about planning their theses and dissertations. This speaker advised students to follow the *KISS principle*: **Keep It Simple, Stupid.** This piece of advice was not given as an insult or meant to be condescending. The speaker understood that there is a natural tendency to want to answer too many questions in one brilliantly conceived and wonderfully designed experiment.

Principle of parsimony
The belief that explanations of phenomena and events should remain simple until the simple explanations are no longer valid.

With a two-group or multiple-group design, there is an obvious limit to how many questions you can ask because of the single IV in those designs. With a factorial design, however, the sky seems to be the limit—you can use as many IVs with as many levels as you wish. Of course, you should always bear in mind that you are complicating matters when you add IVs and levels. Remember the two problems we mentioned earlier in the chapter. One complication occurs in actually conducting the experiment: More participants are required, more experimental sessions are necessary, you have more chances for things to go wrong, and so on. A second complication can occur in your data interpretation: Interactions between four, five, six, or more IVs become nearly impossible to interpret. It is probably this reason that explains why most factorial designs are limited to two or three IVs. Wise detectives limit their investigations to a few leads at a time rather than trying simultaneously to chase down every lead they get.

Variations on Factorial Designs

In Chapters 10 and 11 we saw two variations that have carried with us to this chapter: comparing different amounts of an IV and using measured IVs. For factorial designs we add to this list the use of more than two IVs. Let's review the two carryovers before looking at the new variation.

Comparing Different Amounts of an IV We have already mentioned this variation earlier in the chapter. When we created the hypothetical experiment diagrammed in Figure 12-10, we compared three different types (levels) of customers.

A caution is in order about adding levels to any IV in a factorial design. In the multiple-group design, when you used an additional level of your IV, you added only one group to your experiment. When you add a level to an IV in a factorial design, you add *several* groups to your experiment because each new level must be added under each level of your other independent variable(s). For example, expanding a 2×2 design to a 3×2 design requires 6 groups rather than 4. Enlarging a $2 \times 2 \times 2$ design to a $3 \times 2 \times 2$ design means using 12, rather than 8, groups. Adding levels in a factorial design increases groups in a multiplicative fashion.

Using Measured IVs It will probably not surprise you to learn that we can use nonmanipulated IVs in factorial designs also. It is important to remember that using a measured rather than a manipulated IV results in **ex post facto research**. Without the control that comes from directly causing an IV to vary, we must exercise extreme caution in drawing conclusions from such studies. Still, ex post facto studies give us our only means of studying IVs such as sex or personality traits.

Because factorial designs deal with more than one IV at a time, we can develop an experiment that uses one manipulated IV and one measured IV at the same time. Juan Zapatel, a student at the University of Texas at San Antonio and his faculty sponsor, Stella Garcia-Lopez, (2004) conducted research that provides a good example of just such an experiment. They asked participants to rate song lyrics as a function of the participants' race and the genre of the music.



After reading this one-sentence description of Zapatel and Garcia-Lopez's experiment, can you tell which IV was manipulated and which was measured?

If you decided that participant race was the measured IV and that musical genre was the manipulated IV, you are correct. Of course, participant race is always a measured variable—we cannot “cause” our participants to be different races. Thus, the musical genre *had* to be the manipulated variable. Indeed, Zapatel and Garcia-Lopez labeled the lyrics they showed participants as Black rap, Latin rap, Christian, or alternative. (They actually used only two sets of lyrics that were unfamiliar to participants, with the varying labels.)

What are the implications for interpreting information from an experiment that uses both manipulated and measured IVs? We must still be cautious about interpreting information from the measured IV(s) because we did not cause the levels to vary. On the other hand, we are free to interpret information from manipulated IVs just as usual. Thus, Zapatel and Garcia-Lopez could be more certain about the results from the different labels for the lyrics than they could be about why any effects owing to participant race occurred.

Dealing With More Than Two IVs Designing an experiment with more than two IVs is probably the most important variation of the factorial design. In this section we will discuss the use of factorial designs with three IVs. Larger designs with more than three IVs follow the same basic strategies that we outline here. Again, we must caution you against adding IVs to an experiment without a good reason.

Ex post facto research

A research approach in which the experimenter cannot directly manipulate the IV but can only classify, categorize, or measure the IV because it is predetermined in the participants (e.g., IV = sex).

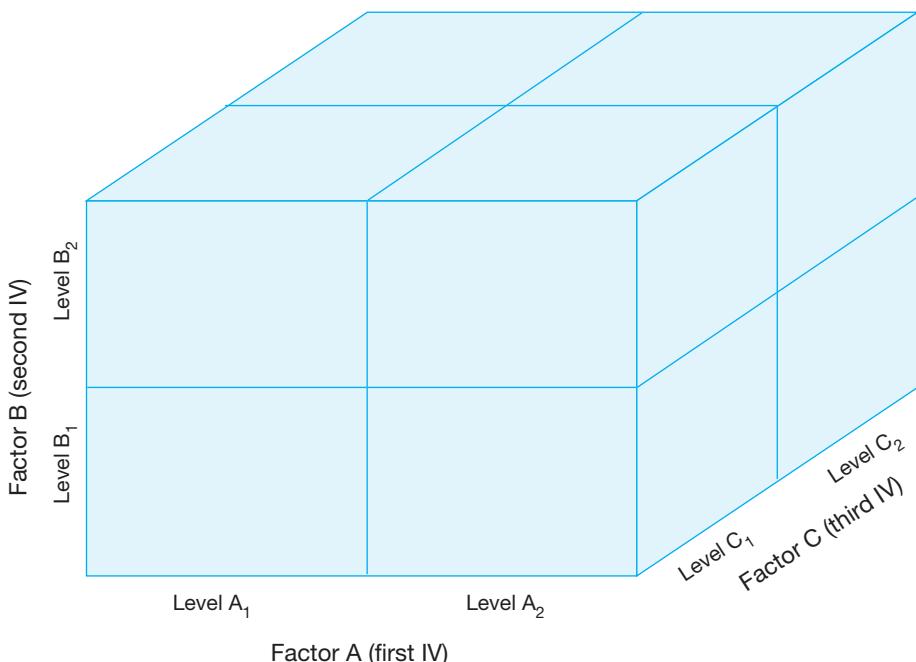


FIGURE 12-11 Simplest Possible Three-Way Factorial Design ($2 \times 2 \times 2$).

Figure 12-11 depicts the simplest possible factorial design with three IVs (often referred to as a **three-way design**). As you can see, it is somewhat difficult to draw three-dimensional designs on a two-dimensional surface. This design has three IVs (A, B, and C), each with two levels. Thus, this design represents a $2 \times 2 \times 2$ experiment. We could expand this design by adding levels to any of the IVs, but remember that the number of treatment combinations increases in a multiplicative rather than an additive fashion.

Three-way design A factorial design with three IVs.

Let's look at a hypothetical example of a three-way design. If you look back at Figure 12-3, you will remember that we conceptualized an extension of McClellan and Woods's (2001) customer hearing study by adding a second IV, the sex of the salesclerks. Imagine that we were also interested in testing for the effects of the customers' sex, in testing whether male or female customers receive help from salesclerks more quickly. This change would transform the design in Figure 12-3 to the design shown in Figure 12-12. We have "exploded" the design so that you can easily see each of the eight treatment combinations. Notice that a group of participants is specified for each of those different treatment combinations.



Figure 12-12 specifies eight different combinations of treatments (the three IVs). Would this design require eight different groups of customers? Why or why not?

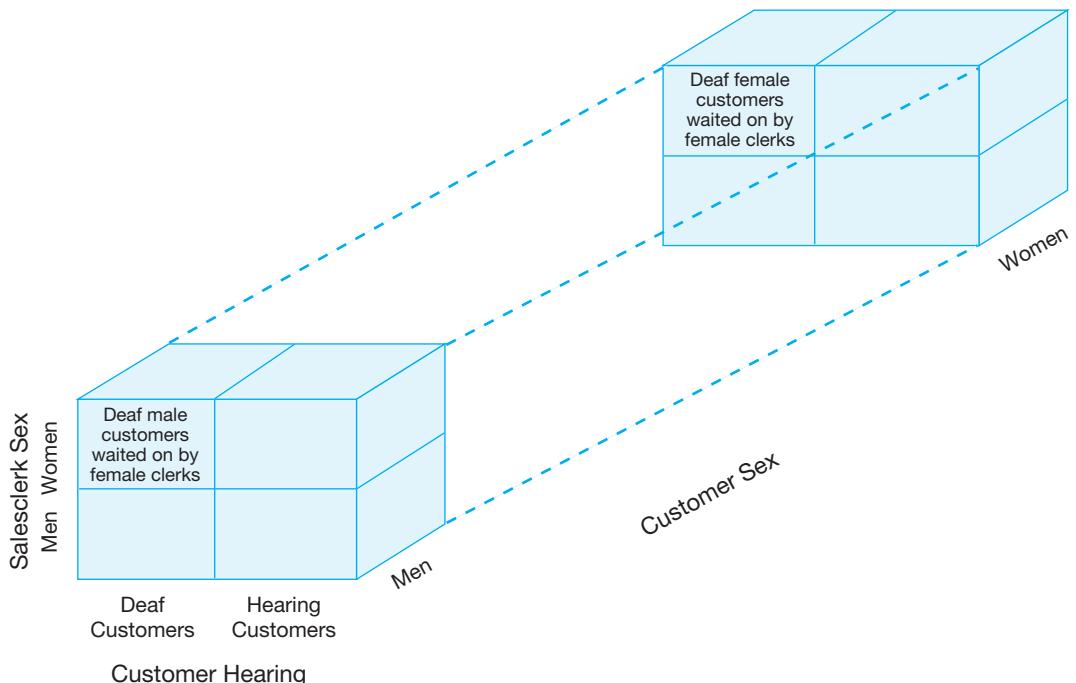


FIGURE 12-12 Second Expansion of McClellan and Woods's (2001) Hearing Disability Study.

This design would require eight different groups of clerks if it is planned as a completely between-groups design. In that case, you would have to create eight different groups of clerks through random assignment. The design would also require eight different groups if you used matched sets of clerks. One possibility that would reduce the number of clerks needed would be the use of repeated measures. Could any of the three IVs be used with repeated measures? Clearly, the sex of the salesclerks is out—a person cannot be both male and female. Customer hearing or sex has the possibility of being used as a repeated measure, depending on how you assign clerks to groups. Salesclerks could wait on both hearing and deaf customers. Likewise, clerks could wait on both male and female customers.

One more consideration is necessary before you could advocate the use of repeated measures: the dependent variable (DV). Can the DV be used more than once with the same salesclerks? If we continue to use response time as the DV, the answer would be yes. There is nothing that would prevent us from measuring a salesclerk's response time to more than one customer. On the other hand, some DVs cannot be used more than once. Again, remember the Burkley et al. (2000) experiment from Chapter 2 in which participants solved anagrams. We cannot use the same anagrams in a repeated-measures situation because the participants would have already solved them. Using repeated measures on anagram solving would require two different but equal sets of anagrams, which should be possible to design.

Thus, the not-so-short answer to the short question in the previous Psychological Detective section is “It depends.” Although the design in Figure 12-12 specifies eight groups of sales-clerks, it does not necessarily specify eight *different* groups of clerks. Our determination of the number of groups depends on whether we use independent groups or repeated measures.

One final point remains for us to cover about factorial designs with more than two IVs. We have mentioned interaction effects several times in this chapter. As we add IVs, we also add more interaction effects to our design. If we used three IVs (A, B, and C), we would obtain information about four interactions: AB, AC, BC, and ABC. To make these interactions more specific, let’s use Figure 12-12 as our example. Given that the experiment diagrammed in Figure 12-12 has customer hearing, salesclerk sex, and customer sex as IVs, our statistical tests will evaluate interactions for customer hearing and clerk sex, customer hearing and customer sex, clerk sex and customer sex, and for all three variables simultaneously (customer hearing, salesclerk sex, and customer sex). You will find as many interaction effects as there are unique combinations of the treatments (see Appendix C).



Imagine that you had an experiment with four IVs: A, B, C, and D. List all possible interactions.

Did you find that there are 11 possible interactions? Let’s look at the possibilities:

Two-way (two IVs) interactions: AB, AC, AD, BC, BD, CD

Three-way (three IVs) interactions: ABC, ABD, ACD, BCD

Four-way (all four IVs) interactions: ABCD

The idea of evaluating 11 different interactions, including one with a possibility of all four variables interacting at once, may be challenging enough for you to understand why we suggest that you limit your experiments to no more than three IVs.

Kelly Early and Amy Holloway, students at Grove City College in Grove City, Pennsylvania, and their faculty sponsor, Gary Welton (2005), studied the factors that influence a person’s willingness to forgive. Based on the previous research that examined various factors in isolation, they combined three factors in their research: intentionality, severity, and friendship.

Severity was manipulated by including either a major or minor altercation. Relationship was manipulated by indicating that the person was either the participant’s closest friend or just an acquaintance. Intentionality was manipulated by describing the transgression as either an intentional offense or an unintentional offense. (Early et al., 2005, p. 55)

Early et al. found all three main effects to be significant. Participants were more likely to forgive a minor altercation, to forgive a best friend, and to forgive an unintentional offense. Although the researchers conducted an experiment with three IVs, the potentially complex three-way interaction did not turn out to be significant.

■ REVIEW SUMMARY

1. Researchers create **factorial designs** by combining two-group or multiple-group designs.
2. We use factorial designs to test the effects of more than one IV on a particular DV at the same time. These designs allow us to test under conditions that are more like the real world than dealing with only one variable at a time.
3. When choosing a particular factorial design, you must consider your experimental questions, issues of control, and practical matters.
4. In a factorial design we can deal with different types or amounts of an IV or with measured IVs, just as we can with a two-group or multiple-group design.
5. Factorial designs may consist of three or more IVs, although the statistical interpretation of such designs becomes more complicated because of an increasing number of interactions.

■ Check Your Progress

1. Why are factorial designs merely combinations of what you learned about in Chapters 10 and 11? Drawing a picture may help.
2. Suppose a friend told you about her $2 \times 4 \times 3$ experimental design. Draw a diagram of this design. Explain its structure.
3. Describe (a) totally between-groups, (b) totally within-groups, and (c) mixed-groups designs. How are they similar? How are they different?
4. Why should your experimental questions be your first consideration in choosing a factorial design?
5. Suppose you wish to test children from two different racial groups. You would be dealing with a(n) _____ IV.
6. Your friend, who plans to take experimental psychology next term, tells you that she is very excited about taking the class because she already has her experiment planned. She wants to test the effects of parental divorce, socioeconomic status, geographical area of residence, parental education, type of preschool attended, and parents' political preference on the sex-role development of children. What advice would you offer this friend?
7. A researcher has conducted a 2×3 experiment. How many interactions will be calculated?
 - a. 1
 - b. 2
 - c. 3
 - d. 6

Statistical Analysis: What Do Your Data Show?

We are certain that you know this fact by now, but we will remind you that experimental design and statistical analysis go hand in hand. You must plan your experiment carefully, choosing the experimental design that best enables you to ask your questions of interest. Having selected your experimental design from the list of choices we are presenting guarantees that you will be able to analyze your data with a standard statistical test. Thus, you will be spared the “experimental fate worse than death”—collecting your data and finding no test with which to analyze the data.

Naming Factorial Designs

In this chapter we have covered designs with more than one IV. Depending on your statistics course, you may not have reached the point of analyzing data from these more complex designs. We analyze factorial designs with the same type of statistical test that we used to analyze multiple-group designs: analysis of variance (ANOVA). As we mentioned in Chapter 11, we must be able to distinguish among the various ANOVA approaches, so we often modify ANOVA with words that refer to the size of the design and the way we assign participants to groups. Labels you may hear that refer to the size of the design include *factorial ANOVA* as a general term or *two-way ANOVA* or *three-way ANOVA* for designs with two or three IVs, respectively. Alternatively, the size of the design may be indicated as X by Y , where X and Y represent the number of levels of the two factors, as we have noted several times in this chapter. For example, labels that describe how researchers assign participants to groups might include *independent groups*, *completely randomized*, *completely between-subjects*, *completely between-groups*, *totally between-subjects*, or *totally between-groups* for designs that use random assignment for all IVs. Designs that use matching or repeated measures may be called *randomized block*, *completely within-subjects*, *completely within-groups*, *totally within-subjects*, or *totally within-groups*. Designs that use a mixture of “between” and “within” assignment procedures may be referred to as *mixed* or *split-plot factorial*. As you can see, the labels for factorial designs can get quite long. Again, if you understand the principles behind the designs, the names are usually not difficult to interpret. For example, you might hear about a *three-way totally between-groups design* or a *two-way mixed ANOVA*.



Can you “decode” the two examples given in the previous sentence? What types of designs do these two cases indicate?

The design indicated by the *three-way totally between-groups* label would include three IVs (three-way) and would use random assignment of participants in all conditions (between groups). *Two-way mixed ANOVA* refers to an experiment with two IVs, one of which uses random assignment and one of which uses a correlated assignment technique (matching or repeated measures). Notice that these descriptions do not give us a full picture of the design because there is no numerical description that would allow us to know how many levels

there are in each IV. To get the fullest amount of information, a label such as a $2 \times 3 \times 2$ completely *between-groups design* is necessary. From this label we know that there are three IVs, with two, three, and two levels, respectively. We also know that the experimenter assigned participants to each of the IVs in a random manner. Notice how much information about an experimental design we can pack into a short, descriptive label.

Planning the Statistical Analysis

In Chapter 11 we featured a hypothetical experiment conducted to investigate the effects of dress on the time it took salesclerks to wait on customers. That example used a multiple-group design because we compared three different styles of clothing to find out how they affected clerks' responses. It should be clear to you that we must derive a new hypothetical experiment to use as an example because we are now dealing with a new design.



How could we make a slight alteration in our clothing style experiment of Chapter 11 in order to make it appropriate for use as an example in this chapter? Give this question some careful thought—make the alteration as simple as possible. Draw a block diagram of your proposed experiment.

As is often the case when you are designing an experiment, there are many possible correct answers. The key feature that your altered experiment must contain is the *addition* of a second IV. By keeping the clothing difference as one IV and adding a second IV, your design should resemble Figure 12-1. Although you could have included more than two levels of either or both of your IVs, that would complicate the design. In asking you to keep your altered design simple, we also assumed you would not choose a three-IV design.

Our statistical example for this chapter builds on our example in Chapter 11. You will remember from Chapter 11 that we compared three different styles of clothing to determine whether the clothes made any difference in how quickly salesclerks responded. Indeed, we found that clerks responded more quickly to customers in dressy or casual clothes than to customers in sloppy clothes. Suppose you are examining the data from the previous experiment and you think you detected an oddity in the data: It appears to you that salesclerks may have responded differently to female and male customers in addition to the different styles of dress. You decide to investigate this question in order to find out whether both customer sex and dress affect salesclerks' response times to customers. Because there was no difference between responses to customers in dressy and casual clothing (see Chapter 11), you decide to use only casual and sloppy clothes. Thus, you have designed a 2×2 experiment (see Figure 12-13) in which the two IVs are clothing style (casual and sloppy) and customer sex (male and female).

Rationale of Factorial ANOVA

The rationale behind ANOVA for factorial designs is basically the same as we saw in Chapter 11, with one major modification. We still use ANOVA to partition (divide) the

		Clothing (Factor A)	
		Casual (A_1)	Sloppy (A_2)
Customer Sex (Factor B)	Female (B_2)		
	Male (B_1)		

FIGURE 12-13 Hypothetical Experiment Using Clothing and Customer Sex IVs.

variability into two sources—**treatment variability** and **error variability**; however, with factorial designs, the sources of treatment variability increase. Instead of having one IV as the sole source of treatment variability, factorial designs have multiple IVs and their interactions as sources of treatment variability. Thus, rather than partitioning the variability as shown in Figure 11-4 (page 247), we would divide the variability as shown in Figure 12-14. The actual distribution of the variance among the factors would depend, of course, on which effects were significant. If you used a factorial design with three IVs, the variability would be partitioned into even more components.

You might guess that we will add statistical formulas because we have added more components to the statistical analysis. You would be correct in this guess. You can turn back to Chapter 11 to review the general ANOVA equations for the one-IV situation. For a two-IV factorial design, we use the following equations:

$$F_A = \frac{\text{IV A variability}}{\text{error variability}} \quad F_B = \frac{\text{IV B variability}}{\text{error variability}} \quad F_{A \times B} = \frac{\text{interaction variability}}{\text{error variability}}$$

These equations allow us to evaluate separately the effects of each of the two IVs as well as their interaction. If we used a larger factorial design, we would end up with an F ratio for each of the IVs and each interaction.

Understanding Interactions

When two variables interact, their joint effect may not be obvious or predictable from examining their separate effects. Let us cite one of the most famous examples of an interaction

Treatment variability

Variability in DV scores due to the effects of the IV (also known as *between-groups variability*).

Error variability

Variability in DV scores due to factors other than the IV, such as individual differences, measurement error, and extraneous variation (also known as *within-groups variability*).

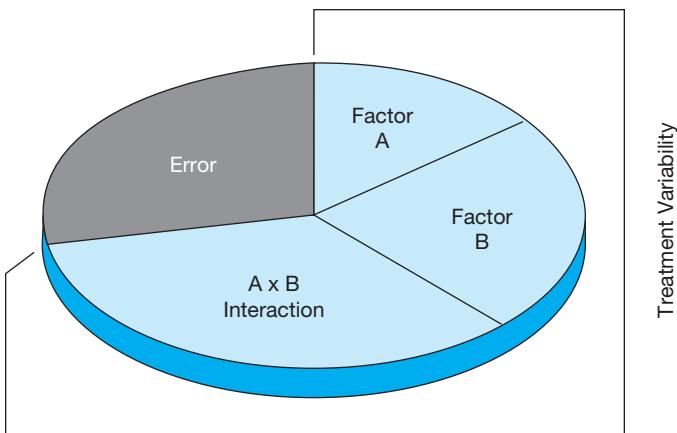


FIGURE 12-14 Partitioning the Variability in a Factorial Design With Two IVs.

effect. Many people find drinking a glass or two of wine to be a pleasurable experience. Many people find taking a drive to be relaxing. What happens if we combine these two activities—do we end up with an extremely pleasurable and relaxing experience? Of course not. We may end up with deadly consequences. Interaction effects often occur with different

drugs. Hence, you often hear very strict guidelines and warnings about combining various drugs. Combinations of drugs, in particular, are likely to have **synergistic effects** so that a joint effect occurs that is not predictable from either drug alone. You may have seen interaction or synergistic effects when two particular children are together. Separately, the children are calm and well behaved; however, when they are put together, watch out! Detectives often face interaction effects in their work. A seemingly normal person, when confronted with a stressful day and an irritating situation, may react violently. Neither the stress nor the annoyance alone would have led to a violent crime, but the combination could be deadly.

Synergistic effects

Dramatic consequences that occur when you combine two or more substances, conditions, or organisms. The effects are greater (or less) than what is individually possible.

Remember our earlier discussion of Figure 12-4 and the interaction pattern found there? In an experimental situation we are concerned with how different levels of different IVs interact with respect to the DV. A significant interaction means that the effects of the various IVs are not straightforward and simple. For this reason we virtually ignore our IV main effects when we find a significant interaction. Sometimes interactions are difficult to interpret, particularly when we have more than two IVs or many levels of an IV. A strategy that often helps us make sense of an interaction is graphing it. By graphing your DV on the y axis and one IV on the x axis, you can depict your other IV with lines on the graph (see Chapter 9). By studying such a graph, you can usually deduce what happened to cause a significant interaction. For example, by examining Figure 12-4 you can see that the results for the customers' disability were not constant based on the sex of the salesclerks. The disability affected the behavior of male clerks to a much greater degree than it affected the behavior of female clerks. Thus, the effects of a customer's hearing disability were not straightforward; they depended on whether female or male salesclerks were being observed. Remember that an interaction is present when the effect of one IV *depends on the specific level of the other IV*.

When you graph a significant interaction, you will often notice that the lines of the graph cross or converge. This pattern is a visual indication that the effects of one IV change as the second IV is varied. Nonsignificant interactions typically show lines that are close to parallel (or somewhat close to parallel), as you saw in Figure 12-8. Riepe (2004) found that education level affected achievement motivation but that gender did not. In addition, there was no interaction between the two IVs (the lines on his graph were nearly parallel). As we cover our statistical examples in the next few pages, we will pay special attention to the interaction effects.

Interpretation: Making Sense of Your Statistics

Our statistical analyses of factorial designs will provide us more information than we got from two-group or multiple-group designs. The analyses are not necessarily more complicated than those we saw in Chapters 10 and 11, but they do provide more information because we have multiple IVs and interaction effects to analyze.

Interpreting Computer Statistical Output

As in Chapters 10 and 11, we will examine generic computer printouts in this section. If you look back at Figure 12-2, you will see that we have three different ANOVAs to cover, based on how we assign the participants to groups. We will deal with 2×2 analyses in these three different categories to fit our clothing-by-customer-sex experiment.

Two-Way ANOVA for Independent Samples The two-way ANOVA for independent samples requires that we have two IVs (clothing style and customer sex) with independent groups. To create this design we would use four different randomly assigned groups of salesclerks, one for each possible combination pictured in Figure 12-13. The DV scores (see Table 12-1) represent clerks' response times in waiting on customers.



Reprinted from *The Chronicle of Higher Education*. By permission of Mischa Richter and Harald Balken.

"On the other hand, if you're not interested in good and evil, this one would give you a good understanding of statistical probability."

A good background in statistics should keep you from becoming this desperate.

TABLE 12-1 Hypothetical Sales Clerks' Response Times (in Seconds) for Comparing Clothing Styles and Sex of Customers

		TYPE OF CLOTHING			
CUSTOMER SEX	Female	Casual	Sloppy	Female $M = 49.83$	
		46 39 50 52 48 54 $M = 48.17$	37 47 44 62 49 70 $M = 51.50$		
Male		38 50 38 44 49 55 $M = 45.67$	47 69 69 74 77 76 $M = 68.67$	Male $M = 57.17$	
		Casual $M = 46.92$	Sloppy $M = 60.08$		
		Total Population $M = 53.50$			

Computer Results The descriptive statistics appear at the top of Table 12-2. You can see that salesclerks took almost a minute on average to wait on the customers: The mean for the “total population” (all 24 clerks) was 53.50 seconds. The means for the clerks waiting on customers in casual and in sloppy clothes were 46.92 and 60.08 seconds, respectively; for the female customers and the male customers, the means were 49.83 and 57.17 seconds, respectively. The last set of descriptive statistics shows the combination of the two clothing styles and the two sexes. Clerks waiting on casually dressed women averaged 48.17 seconds; those waiting on casually dressed men averaged 45.67 seconds. For the sloppy clothing, clerks waiting on women had a mean of 51.50 seconds, whereas the clerks waiting on men took an average of 68.67 seconds. Again, to make sure you entered the data correctly, you could check these means with a hand calculator in a matter of minutes.

Source table A table containing the results of ANOVA. *Source* refers to the source of different types of variation.

The **source table** for the completely randomized factorial design appears at the bottom of Table 12-2. In the body of the source table, we want to examine only the effects of the two IVs (clothing and customer sex) and their interaction. The remaining source (w. cell or Within) is the error term and is the term we use to test the IV effects. One important item to note is that different programs are likely to label the error term with a variety of different names, such as “W. Cell” in this case, “error,” etc. (If you need to review concepts like the sum of squares or mean squares, refer back to Chapter 11.) When we examine the main effects, we find that “Clothes” produced an F ratio of 11.92, with a probability of occurring by chance of .003 (from rounding). The effect of customer sex shows an F ratio of 3.70, with a probability of .07. Can you verify these probabilities in the F table in the back of the book? You should find that the probability of “Clothes” falls below the .01 level in the table. The sex IV shows **marginal significance**, which we usually attribute to probabilities of chance

TABLE 12-2 Computer Output for Two-Way ANOVA for Independent Samples

TABLE OF MEANS (IN SECONDS):

		CLOTHES			
		CASUAL	SLOPPY	ROW M	
Customer	Female	48.17	51.50	49.83	
Sex	Male	45.67	68.67	57.17	
	Column M	46.92	60.08		
		Pop. M = 53.50			
SOURCE TABLE					
SOURCE	SS	df	MS	F	p
Clothes	1040.17	1	1040.17	11.922	.0025
Customer Sex	322.67	1	322.67	3.698	.068
Clo × Cust Sex	580.17	1	580.17	6.649	.0179
W. Cell	1745.00	20	87.25		
Total	3688.00	23			

between 5% and 10%. Although marginal significance is not within the normal significance range, it is close enough that many experimenters discuss such results anyway. Keep in mind that when you deal with higher and higher probabilities of chance you are taking a greater risk of making a **Type I error** (see Chapter 9).

Our next step is to examine the interaction between the two IVs. Notice that we have one two-way interaction because we have just the two IVs. The interaction between clothing and customer sex produced an *F* ratio of 6.65 and has *p* = .02, therefore denoting significance. Remember that a significant interaction renders the main effects moot because those main effects are qualified by the interaction and are not straightforward. Thus, to make sense out of these results, we must interpret the interaction. The first step in interpreting an interaction, as we saw, is to draw a graph of the results from the descriptive statistics. Figure 12-15 depicts this interaction.

Translating Statistics Into Words Remember that we are justified in drawing conclusions from our statistics *only if* we are certain that our experimental design and procedures had sufficient control to eliminate extraneous variables. Computers and statistical tests work only with the data you provide; they are not able to detect data from flawed experiments.

To interpret our statistics, let's return to the graph of our significant interaction (Figure 12-15).

Marginal significance

Refers to statistical results with a probability of chance between 5% and 10%; in other words, almost significant, but not quite. Researchers often talk about such results as if they reached the *p* = .05 level.

Type I error Accepting the experimental hypothesis when the null hypothesis is true.



Study Figure 12-15 carefully. What do you think causes the significant interaction? In other words, why do the results of one IV depend on the particular level of the second IV?

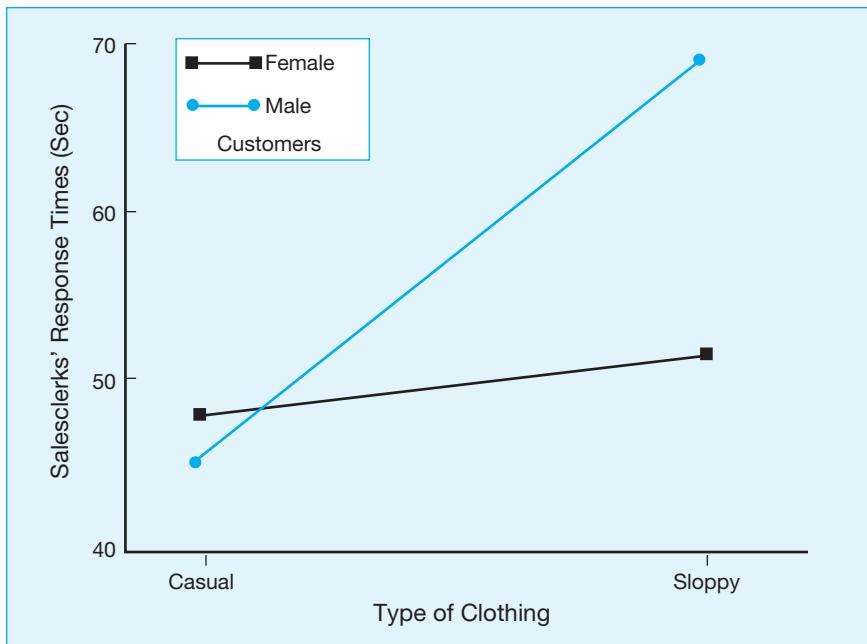


FIGURE 12-15 Interaction for Clothing and Customer Sex Experiment.

You should remember that interactions occur when the lines on a graph cross or converge; parallel lines indicate no interaction. The crossing lines shown in Figure 12-15, in conjunction with the low probability of chance for the interaction term, denote a significant interaction. When we examine the figure, the point that seems to differ most from the others represents the clerks' response times to male customers in sloppy clothes. This mean is considerably higher than the others. Thus, we would conclude that clerks take longer to wait on men who are sloppily dressed than on other customers. Notice that our explanation of an interaction effect *must* include a reference to both IVs in order to make sense—the previous sentence mentions the *male* customers in *sloppy* clothes.



Look at Figure 12-15. Can you figure out why this interaction qualifies the significant and marginally significant main effects?

If we had drawn conclusions from the main effects, we would have decided that clerks waiting on customers in casual clothes were faster (46.92 seconds) than clerks waiting on those in shabby clothes (60.08). Also, we would have concluded that clerks waiting on men took marginally longer (57.17) than those waiting on women (49.83). When you look at Figure 12-15, does it appear that clerks actually waited on sloppily dressed customers more

slowly than on those customers dressed in casual clothes? No, only clerks waiting on *men* in sloppy clothes had longer response times. Does it seem that men received assistance more slowly than women? No, clerks responded slowly only to the men in *sloppy clothes*. If you attempt to interpret the main effects in a straightforward fashion when you have a significant interaction, you end up trying to make a gray situation into a black-and-white picture. In other words, you would be guilty of oversimplifying the results. Remember that an interaction occurs when the effect of one IV depends on the *specific* level of the other IV.

The final step in interpreting the results is communicating our results to others. You will remember that we use a combination of statistical results and words to convey our findings in APA style. Here is one way you could present the results from this experiment:

The effect of the clothing on clerks' response times was significant, $F(1, 20) = 11.92, p = .003$. The customer sex effect was marginally significant, $F(1, 20) = 3.70, p = .069$. The main effects were qualified by a significant interaction between clothing and customer sex, $F(1, 20) = 6.65, p = .018, \eta^2 = 0.25$. The results of the interaction appear in Figure 1 [see Figure 12-15]. Visual inspection of the graph shows that clerks' response times for the sloppy clothes male customer condition were higher than those for the other conditions.

Once again, our goal is to communicate our results clearly with words alone. Although the concept of interactions is somewhat complicated for readers with no statistical background, we hope that anyone could understand our explanation of the results. To understand the interaction fully, however, you should conduct further statistical tests. These tests are somewhat advanced—we suggest that you consult with your advisor if your research involves a significant interaction. (Be sure to remember this warning because you will see significant interactions again in this chapter.) Also, we remind you that η^2 is a measure of effect size (see Chapter 10). We included an effect-size estimate only for the interaction because the significant interaction supersedes the main effects.

We want you to note two important points from this presentation of results. First, compared to the examples presented in Chapters 10 and 11, this written summary is longer. When you deal with factorial designs, you have more results to communicate, and your presentations will be longer. Second, although we presented the significance of the interaction and referred to a figure, we did not fully interpret the interaction. Results sections in experimental reports are only for presenting the results, not for full interpretation. As the APA's *Publication Manual* (2001) notes in describing the results section, "discussing the implications of the results is not appropriate here" (p. 20).

Two-Way ANOVA for Correlated Samples The two-way ANOVA for correlated samples requires that we have two IVs (here, clothing and customer sex) with correlated groups for both IVs. Most often, researchers would form these correlated groups by matching or by using repeated measures. In our example of the clothing–customer sex experiment, it is possible to use repeated measures on both IVs. We would merely get one sample of salesclerks and have them wait on customers of both sexes wearing each style of clothing.



Can you detect a flaw in the logic of having a single group of clerks waiting on all four types of customers?

There is no obvious flaw in using repeated measures in this experiment. Salesclerks are certainly used to waiting on a variety of different customers in a short time span, so there should not be a problem in creating such a condition for this experiment. On the other hand, there are many experimental situations in which it would not be appropriate to use repeated measures. For example, imagine a classroom experiment that involved students reading chapters on the same topic from two different textbooks and then taking a quiz on each chapter. This hypothetical experiment would result in an advantage for the second quiz on each topic. Specifically, when students took that quiz, they would have read *two* chapters about the same topic. On the other hand, when they took their first quiz on each topic, they would have read only one chapter on the topic. Clearly, repeated measures on this textbook variable would not be appropriate. Thus, we would have to form correlated groups by forming matched groups of students. It is critical that you examine your experimental question and procedures for assigning participants to the groups in order to determine which specific experimental design to use.

Computer Results The results for the factorial repeated-measures analysis appear in Table 12-3. You will note that the group means are identical to those in Table 12-2 (the ANOVA for independent samples). This result is to be expected; we are analyzing the same numbers from Table 12-1. Because we used a different strategy for composing the groups in the analysis, however, our ANOVA results should be different.

The source table for the two-way ANOVA with correlated groups appears in the bottom portion of Table 12-3. The clothing effect is significant at the .001 level, and the sex effect is significant at the .014 level. Both of those effects, however, are qualified by the significant clothing-by-sex interaction ($p = 0.001$). Remember that this interaction effect signifies that the results of the IVs are not consistent across each other. To make sense of the interaction, we must plot the means for the combinations of clothing and customer sex. This interaction appeared in Figure 12-15.

TABLE 12-3 Computer Output for Two-Way ANOVA for Correlated Samples

TABLE OF MEANS (IN SECONDS):

Clothes

		CASUAL	SLOPPY	ROW M
Customer	Female	48.17	51.50	49.83
Sex	Male	45.67	68.67	57.17
Column M		46.92	60.08	

Pop. M = 53.50

SOURCE TABLE

Source	SS	df	MS	F	p
Clothes	1040.17	1	1040.17	24.688	.001
Customer Sex	322.67	1	322.67	7.658	.014
Clo × Cust Sex	580.17	1	580.17	13.770	.001
Residual	632.00	15	42.13		

Translating Statistics Into Words One important lesson to learn from these different analyses of the same data deals with the power of designing experiments in different ways. For example, take a minute to compare the source table for this analysis (Table 12-3) to the source table for the completely randomized analysis (Table 12-2). You will notice in the correlated-samples design that the F ratios are larger and the probabilities are smaller for both the IVs and their interaction. In the previous chapters we told you that using correlated samples reduces error variability by reducing some of the between-subjects variability. The result typically is a stronger, more powerful test of the treatment effects, as is shown in this case.

To interpret the results of this experiment fully, we must explain the interaction shown in Figure 12-15. We will be briefer at this point because we have already carefully examined the interaction in the previous analysis section. Again, it is clear that the interaction occurred because salesclerks were slower to respond to sloppily dressed male customers than other customers; however, the effect was specific to that type of clothing on that sex of customer. Salesclerks were not slower in waiting on men in general because the time to wait on casually dressed men was not higher. By the same token, clerks were not generally slower in waiting on sloppily dressed customers because they waited on sloppily dressed women in about the same time as casually dressed customers. Thus, we must confine our conclusion to slower responses to one mode of dress for one sex of customer. Notice that explaining an interaction forces us to refer to both IVs in the same explanatory sentence—we can't ignore one IV to focus on the other.

Of course, we must still communicate our results to other parties using APA style. We rely on our standard combination of words and numbers for the summary of the results—words to explain and numbers to document the findings. One possible way of summarizing these results follows:

Both the main effects of clothing and customer sex were significant, $F(1, 5) = 24.69, p = .001$ and $F(1, 5) = 7.66, p = .014$, respectively. Furthermore, the interaction of clothing and customer sex was also significant, $F(1, 5) = 13.77, p = .001, \eta^2 = .48$. This interaction appears in Figure 1 [see Figure 12-15]. Salesclerks waiting on sloppily attired male customers were considerably slower than clerks with any other combination of customer sex and clothing.

You would provide a fuller explanation and interpretation of this interaction in the discussion section of your experimental report.

Two-Way ANOVA for Mixed Samples The two-way ANOVA for mixed samples requires that we have two IVs (here, clothing and customer sex) with independent groups for one IV and correlated groups for the second IV. One possible way to create this design in our clothing–customer sex experiment would be to use a different randomly assigned group of salesclerks for each customer sex. Clerks waiting on each sex, however, would assist customers attired in both types of clothing. Thus, customer sex would be independent groups and constitute a between-subjects variable, whereas clothing would use repeated measures and be a within-subjects variable. Looking at Figure 12-13, note that different salesclerks would wait on men or women (looking down the diagram) and would wait on customers in both casual *and* sloppy clothes (across the top of the diagram). The DV scores (see Table 12-1) still represent response times of clerks to customers, but the same clerks produced the response times for each clothing type within the two customer sexes. This design is efficient because it requires fewer clerks to conduct the study, and it minimizes the individual differences within the response times.



Could you reverse the between- and within-subjects variables in this design?

This question basically asks whether you could have clerks wait on either men and women in casual clothing or men and women in sloppy clothing. In this case the answer is yes—there is nothing magical about waiting on customers of the same sex or customers dressed in the same type of clothing. Of course, it is not always possible to “reverse” a design as we have in this example. Remember that we alerted you earlier in the text that not all IVs can be treated as within-subject variables.

Computer Results The descriptive statistics appear at the top of Table 12-4. Once again, the descriptive statistics did not change from our first and second analysis; we are still analyzing the same data (see Table 12-1) for demonstration purposes.

The source table appears at the bottom of Table 12-4. As you can see from the headings, the between-subjects effects (independent groups) and the within-subjects effects (repeated measures) are divided in the source table. This division is necessary because the between-subjects effects and within-subjects effects use different error terms. The interaction appears in the within-subjects portion of the table because it involves repeated measures across one of the variables involved.

TABLE 12-4 Computer Output for Two-Way ANOVA for Mixed Samples

TABLE OF MEANS (IN SECONDS):

CLOTHES					
		CASUAL	SLOPPY	ROW M	
Customer	Female	48.17	51.50	49.83	
Sex	Male	45.67	68.67	57.17	
	Column M	46.92	60.08		Pop. M = 53.50

SOURCE TABLE						
SOURCE	SS	df	MS	F	p	
<i>Between-subjects effects</i>						
Cust Sex	322.67	1	322.67	2.422	.151	
Error (Cust Sex)	1332.33	10	133.23			
<i>Within-subjects effects</i>						
Clothes	1040.17	1	1040.17	25.206	.001	
C to X Cust Sex	580.17	1	580.17	14.059	.004	
Error	412.67	10	41.27			



Which IV is the between-subjects variable and why? Which IV is the within-subjects variable and why?

No, that wasn't a trick question—it's just a simple review query to make sure you're paying attention. Customer sex is the between-subjects variable because different salesclerks waited on men or women. The clothing is the within-subjects variable because each salesclerk waited on customers in both sloppy and casual clothes (repeated measures).

The information for the customer sex IV showed an F ratio of 2.42 and a probability of chance of .15; therefore, the customers' sex made no significant difference in the salesclerks' response times. The clothing effect yielded an F ratio of 25.21 with a probability of .001, a significant finding. We also notice that the interaction of clothing and customer sex is significant ($p = .004$), with an F ratio of 14.06. Because of this significant interaction, we would not interpret the significant clothing result. Once again, we must graph the interaction in order to make sense of it (see Figure 12-15).

This set of ANOVA results differs from either of the other two analyses in this chapter, again demonstrating the importance of experimental design in determining significance. Interestingly, the customer sex effect was the weakest in this design. Kirk (1968) noted that tests of between-subjects factors are relatively weaker than tests of within-subjects factors in a mixed (split-plot) design. The results of this analysis, when compared to the previous analysis, demonstrate that point.

Translating Statistics Into Words We have already completed some of our interpretation in our coverage of the statistical results. We know that the customer sex effect was not significant and that the clothing variable was significant; however, because the interaction between customer sex and clothing was significant, we ignore the clothing results and interpret the interaction.

Figure 12-15 shows that clerks who waited on casually dressed women, casually dressed men, and sloppily dressed women responded more quickly than those who waited on sloppily dressed men. How can we communicate these findings in APA format? Here's one possibility:

Results from the mixed factorial ANOVA showed no effect of the customer sex, $F(1, 10) = 2.42$, $p = .15$. The clothing effect was significant, $F(1, 10) = 25.21$, $p = .001$. This main effect, however, was qualified by a significant customer-sex-by-clothing interaction, $F(1, 10) = 14.06$, $p = .004$, $\eta^2 = .58$, which appears in Figure 1 [see Figure 12-15]. This interaction shows that salesclerks who waited on sloppily dressed male customers were slower in responding than clerks who waited on casually dressed men or women dressed in either manner.



Note the similarity for the clothing and interaction effects in this mixed analysis and the completely within-groups analysis in the previous section. Why do you think the findings are similar in each case?

To answer this difficult question, you have to be able to see the similarities between these last two analyses. Although they may seem rather different, there is one important similarity between them. Both of these analyses, because of the underlying experimental designs, treat the clothing and interaction effects as within-groups effects. Thus, the two ANOVAs are essentially analyzing the same data in the same manner.

As a final word about the analyses in this chapter, please remember that you cannot analyze the same data using several different analyses in the real world of experimentation. We have used the same data for instructional purposes *and* to demonstrate how you could take one experimental idea and put it in the context of several possible experimental designs.

A Final Note

For simplicity's sake, all the analyses we have shown in this chapter have dealt with IVs with only two levels. You remember from Chapter 11 that we often wish to test IVs with more than two levels. In Figure 12-9 we showed you an example of a 3×3 design. That design had two IVs, each with three levels. Do you remember what happened in Chapter 11 when we found significance for an IV with three levels? To determine what caused the significant findings, we carried out **post hoc comparisons**. These tests that we calculated after finding a significant IV allowed us to determine which levels of that IV differed significantly.

We hope that this issue has occurred to you at some point during this chapter. What should you do, in a factorial design, if an IV with more than two levels turns out to be significant? Assuming that this main effect is not qualified by an interaction, you must calculate a set of post hoc tests to determine exactly where the significance of that IV occurred.

Post hoc comparisons

Statistical comparisons made between group means after finding a significant F ratio.

The Continuing Research Problem

In Chapters 9 and 10 we began this discussion about our hypothetical continuing research problem. Our early interest was in determining whether customers in dressy or sloppy clothing received faster assistance from salesclerks. When we found that dressy clothes were associated with faster responses, we enlarged our research question to include three different types of clothing (Chapter 11). Evidence showed that customers dressed in either casual or dressy clothes were helped by salesclerks more quickly than customers dressed in shabby clothing.

As you saw in Chapter 10, we began our research problem with a fairly simple question, and we got a fairly simple answer. That research led to a slightly more complex question, and so on. You should expect your research problems to show a similar pattern. Although you may start with a simple question and expect one experiment to provide all the answers, that will rarely be the case. Keep your eyes open for new questions that arise after the experiment. Pursuing a line of **programmatic research** is challenging, invigorating, and interesting. Remember that pursuing such a line of research is how most famous psychologists have made names for themselves.

Programmatic research

A series of research experiments concerning a related topic or question.

This chapter was more complex than either Chapter 10 or Chapter 11 because of the extra choice in experimental design. Let's review the steps we took in designing the experiments in this chapter. You may wish to refer to Figure 12-1 to follow each specific question.

1. After conducting our preliminary research in Chapters 10 and 11, we decided to use two IVs (clothing and customer sex) in these experiments. Each IV had two levels (clothing → casual, sloppy; customer sex → men, women). This design allows us to determine the effect of clothing, the effect of customer sex, and the interaction between clothing and customer sex.
2. The DV was the time it took salesclerks to respond to customers.
- 3a. With large numbers of clerks, we randomly formed four groups of clerks, each of which waited on one sex of customer in one type of clothing, resulting in a factorial between-groups design. We analyzed the response times using a factorial ANOVA for independent groups and found that clerks were slower to wait on male customers dressed in sloppy clothing than all other customers (see Table 12-2 and Figure 12-15).
- 3b. In a hypothetical situation with fewer clerks for the experiment, we used repeated measures on both IVs; that is, each salesclerk waited on both sexes of customers attired in both types of clothing, so each clerk waited on four different customers. Thus, this experiment used a factorial within-groups design. We analyzed the data with a factorial ANOVA for correlated groups and found that clerks were slowest in waiting on sloppily dressed men (see Table 12-3 and Figure 12-15).
- 3c. In a third hypothetical situation we randomly assigned salesclerks to the two customer sex groups but used repeated measures on the clothing IV so that clerks waited either on men in both types of clothing or women in both types of clothing. This arrangement resulted in a factorial mixed-groups design (one IV using independent groups, one IV using correlated groups). We analyzed the response times with a factorial ANOVA for mixed groups and found the slowest response times for clerks to male customers dressed in sloppy clothes (see Table 12-4 and Figure 12-15).
4. We concluded that clothing and customer sex interacted to affect salesclerks' response times. Women received help quickly regardless of their attire, but men received help quickly only if they were not sloppily dressed. Men attired in sloppy clothes had to wait longer for help than participants in the other three groups.

■ REVIEW SUMMARY

1. When you use an experimental design with two or more IVs and have only independent groups of participants, the proper statistical analysis is a factorial ANOVA for independent groups.
2. A factorial ANOVA for correlated groups is appropriate when your experimental design has two or more IVs and you used matched groups or repeated measures for all IVs.
3. If you have a design with two or more IVs and a mixture of independent and correlated groups for those IVs, you would use a factorial ANOVA for mixed groups to analyze your data.

4. ANOVA partitions the variability in the DV into separate sources for all IVs and their interactions and for error. F ratios show the ratio of the variation for the experimental effects to the error variation.
5. A significant F ratio for a main effect indicates that the particular IV caused a significant difference in the DV scores.
6. A significant interaction F ratio indicates that the two (or more) IVs involved had an inter-related effect on the DV scores. To make sense of an interaction, you should graph the DV scores.
7. We use APA format to communicate our statistical results clearly and concisely. Proper format includes a written explanation of the findings documented with statistical results.

Check Your Progress

1. You wish to compare the ACT or SAT scores of the freshmen, sophomore, junior, and senior classes at your school as a function of sex. Draw a block diagram of this design. What design and statistical test would you use for this project?
2. You wonder whether test-taking practice and study courses can actually affect SAT or ACT scores. You recruit one group of students to help you. They take the test three times. Then you give them a study course for the test. They take an alternative form of the same test, also three times. Thus, each student has taken two different tests three times (to study practice effects), and each student has taken the study course (to assess its effects). Draw a block diagram of this design. What design and statistical test would you use for this project?
3. You are interested in the same question as in Problem 2, but you recruit two groups of students to help you. One group takes the SAT or ACT three times; the other group has a study course and then takes the SAT or ACT three times. Draw a block diagram of this design. What design and statistical test would you use for this project?
4. What is an interaction effect? Why does a significant interaction render its associated main effects uninterpretable?
5. Suppose you are reading an experimental report. What would you know from the following sentence?

Reading speed was affected by both print size and age, with younger participants reading large print faster and older participants reading small print faster.

- a. The interaction of print size and age was significant.
- b. The interaction of print size and age was not significant.
- c. The interaction of print size, age, and reading speed was significant.
- d. The main effects of print size and age were significant.

6. You wish to determine whether people's moods differ during the four seasons and by sex. What experimental design would you use for this research project? Why?

7. You choose to test people's preferences for fast-food hamburgers in three different restaurants (McDonald's, Burger King, Wendy's) among three different age groups (children ages 4–12, college students, senior citizens). What experimental design would you use for this project? Why?

■ Key Terms

Factors, 259	Mixed assignment, 268	Source table, 288
Independent variables, 259	Principle of parsimony, 277	Marginal significance, 289
Factorial design, 261	Ex post facto research, 278	Type I error, 289
Main effect, 264	Three-way design, 279	Post hoc
Interaction, 264	Treatment variability, 285	comparisons, 296
Independent groups, 268	Error variability, 285	Programmatic
Correlated groups, 268	Synergistic effects, 286	research, 296

■ Looking Ahead

In this chapter we have learned about the most sophisticated of all experimental designs; factorial designs with multiple IVs. This material completes a three-chapter discussion of experimental design, data analysis, and interpretation. In the next chapter we will look at some alternative research approaches that may be useful for you sometime.

Alternative Research Designs

CHAPTER
13

Protecting Internal Validity Revisited

- Examining Your Experiment From the Inside
- Protecting Internal Validity With Research Designs
- Conclusion

Single-Case Experimental Designs

- History of Single-Case Experimental Designs
- Uses of Single-Case Experimental Designs

- General Procedures of Single-Case Experimental Designs
- Statistics and Single-Case Experimental Designs
- Representative Single-Case Experimental Designs

Quasi-Experimental Designs

- History of Quasi-Experimental Designs
- Uses of Quasi-Experimental Designs
- Representative Quasi-Experimental Designs

Internal validity A type of evaluation of your experiment; it asks whether your IV is the only possible explanation of the results shown for your DV.

Confounding Caused by an uncontrolled extraneous variable that varies systematically with the IV.

Extraneous variables Uncontrolled variables that may unintentionally influence the dependent variable (DV) and thus invalidate an experiment.

Cause-and-effect relation Occurs when we know that a particular IV (cause) leads to specific changes in a DV (effect).

Although you may think that by now we have covered every conceivable type of research design that psychologists might use to gather data, you would be wrong. There are many other types of research designs. In this chapter we will consider some designs developed by researchers with specific purposes in mind. We will look first at research designs that protect the internal validity of our experiments.

Protecting Internal Validity Revisited

In Chapter 8 we introduced the concept of **internal validity**. The issue of internal validity revolves around **confounding** and **extraneous variables**. When you have an internally valid experiment, you are reasonably certain that your independent variable (IV) is responsible for the changes you observed in your dependent variable (DV). You have established a **cause-and-effect relation**, knowing that the IV *caused* the change in the DV. For example, after many years of painstaking research, medical scientists know that cigarette smoking *causes* lung cancer. Although there are other variables that can trigger cancer, we know that smoking is a causative agent. Our goal as experimenters is to establish similar cause-and-effect relations in psychology. Experiments that are internally valid allow us to make statements such as “X causes Y to occur” with confidence.

Examining Your Experiment from the Inside

In Chapter 6 we talked about the necessity for controlling extraneous variables in order to reach a clear-cut conclusion from our experiment. It is only when we have designed our experiment in such a way as to avoid the effects of potential extraneous variables that we can feel comfortable about making a cause-and-effect statement; that is, saying that Variable X (our IV) *caused* the change we observed

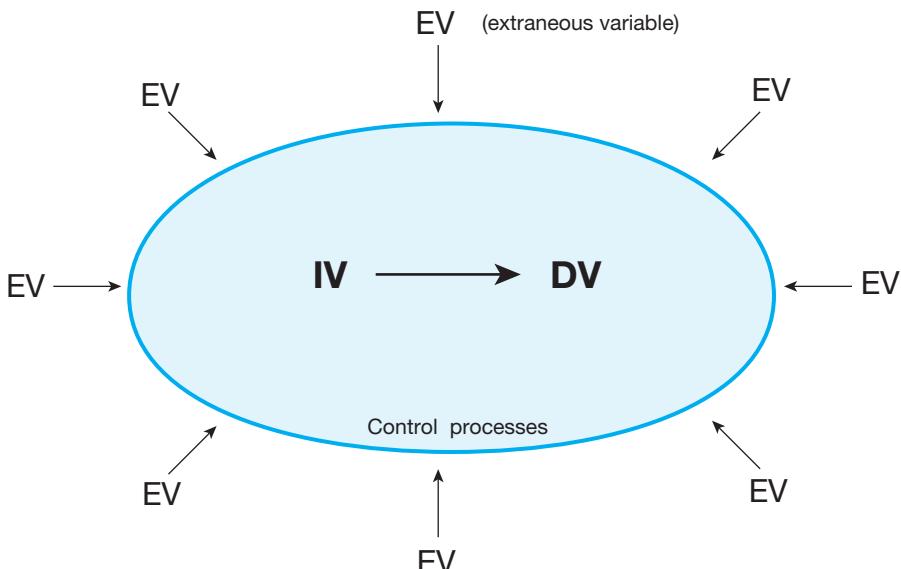


FIGURE 13-1 The Role of Control Processes in Preventing Experimental Confounding.

in Variable Y (our DV). What we are trying to accomplish through our control techniques is to set up a buffer for our IV and DV so that they will not be affected by other variables. This reminds us of a cartoonlike toothpaste commercial we saw—perhaps you have seen it also. When the teeth brushed themselves with the particular brand of toothpaste being advertised, they developed a protective “invisible barrier” against tooth decay. In an analogous manner, our controls give our experiment a barrier against confounding (see Figure 13-1). Similarly, police detectives strive to make their case against a particular suspect airtight. If they have carried out their investigations well, the case against the accused should hold up in court.

Dealing with the internal validity of an experiment is an interesting process. We take many precautions aimed at increasing internal validity as we design and set up our experiment, and we usually evaluate our experiment with regard to internal validity after we have completed the research. If this approach seems a little strange to you, don’t be alarmed—it does seem odd at first. Internal validity revolves around the question of whether your IV actually created any change observed in your DV. As you can see in Figure 13-1, if you learned your lessons from Chapter 8 well and used adequate control techniques, your experiment should be free from confounding and you can indeed conclude that your IV caused the change in your DV. Let’s review briefly.



Imagine you have been given responsibility for conducting the famous Crest test—you are supposed to determine whether brushing with Crest actually does reduce cavities. Your boss wants you to use an experimental group (Crest) and a control group (Brand X) in the experiment. Write down at least five potential extraneous variables for this experiment before reading further.

Were you able to list five possible extraneous variables? The list could be quite long; you may have thought of some possibilities that we didn't. (Although this exercise could easily have been an example dealing with confounding found in Chapter 8, it is also relevant to the issue of internal validity. If you fail to control an important extraneous variable, your experiment will not have internal validity.) Remember, any factor that systematically differs between the two groups (other than the type of toothpaste) could be an extraneous variable that could make it impossible to draw a definite conclusion about the effect of the toothpastes. Here's our (partial) list of possibilities:

- number of times brushed per day
- amount of time spent in brushing per day
- how soon brushing occurs after meals
- types of foods eaten
- type of toothbrush used
- dental genetics inherited from parents
- degree of dental care received
- different dentists' "operational definition" of what constitutes a cavity
- whether the city's water is fluoridated

As we said, this list is not meant to be exhaustive—it merely gives you some ideas of factors that could be extraneous variables. To make certain you understand how an extraneous variable can undermine an experiment's internal validity, let's use an example from the previous list. In addition, we will discover why we take precautions aimed at internal validity *before* the experiment and assess the internal validity of an experiment *afterward*.

When you design the study, you want to make sure that people in the experimental and control groups brush their teeth an equivalent number of times per day. Thus, you would instruct the parents to have their children brush after each meal. Your goal is to have all children brush three times a day. Suppose that you conducted the experiment and gathered your data. When you analyzed the data, you found that the experimental group (Crest) had significantly fewer cavities than the control group (Brand X). Your conclusion seems straightforward at this point: Brushing with Crest reduces cavities compared to brushing with Brand X. As you dig deeper into your data, however, you look at the questionnaire completed by the parents and discover that the children in the experimental group averaged 2.72 brushings a day compared to 1.98 times per day for the children in the control group. Now it is obvious that your two groups differ on two factors: the type of toothpaste used and the number of brushings per day. Which factor is responsible for the lower number of cavities in the experimental group? It is impossible to tell! There is no statistical test that can separate these two confounded factors. You attempted to control the brushing factor before the experiment to assure internal validity, but you could not assess your control technique until after the experiment, when you found out that your experiment was not internally valid. A word to the wise should be sufficient: Good experimental control leads to internally valid experiments.

Remember that we listed nine threats to internal validity in Chapter 8. We also provided you with a variety of control strategies to deal with those threats. Now that you are familiar with research designs, we can explain how some research design strategies eliminate threats to internal validity. As you read about these strategies, you will see that some are a part of those designs discussed in Chapters 10 through 12.

Protecting Internal Validity With Research Designs

There are two approaches you could take to fight the various threats to internal validity. In the first approach you would attempt to come up with nine different answers, one for each threat. Although this approach would be effective in controlling the threats, it would be time consuming and, perhaps, difficult to institute that many different controls simultaneously. Perhaps the idea of controlling the threats through research design occurred to you, even if you could not come up with a specific recommendation. Detectives use standard police procedures to help them protect their cases; experimental design procedures can help us as psychological detectives.

In the three previous chapters we presented you with a variety of experimental designs, often noting various control aspects of those designs; however, we never mentioned the nine general threats to internal validity until this chapter. Can we apply experimental design to these problems? According to Campbell (1957) and Campbell and Stanley (1966), the answer is "yes." Let's take a look at their recommendations.

Random Assignment Although **random assignment** is not a specific experimental design, it is a technique that we can use within our experimental designs. Remember, with random assignment (see Chapter 4) we distribute the experimental participants into our various groups on a random (nonsystematic) basis. Thus, all participants have an equal chance of being assigned to *any* of our treatment groups. The purpose behind random assignment is to create different groups that are equal before beginning our experiment. According to Campbell and Stanley (1966), "[T]he most adequate all-purpose assurance of lack of initial biases between groups is randomization" (p. 25). Thus, random assignment can be a powerful tool. The only drawback to random assignment is that we cannot *guarantee* equality through its use.

Random assignment

This control technique ensures that each participant has an equal chance of being assigned to any group in an experiment.

One caution is in order at this point. Because *random* is a frequently used term when dealing with experimental design issues, it sometimes has slightly different meanings. For example, in Chapters 10 through 12 we repeatedly referred to *independent groups* to describe groups of participants that were not correlated in any way (through matching, repeated measures, or natural pairs or sets). It is not unusual to see or hear such independent groups referred to as *random groups*. Although this label makes sense because the groups are unrelated, it is also somewhat misleading. Remember in Chapter 10 when we first talked about matching participants? At that point we stressed that after making your matched pairs of participants, you *randomly assigned* one member of each pair to each group. The same is true of naturally occurring pairs (or sets) of participants. These randomly assigned groups would clearly *not* be independent. Because of the power of random assignment to equate our groups, we should use it at every opportunity. Campbell and Stanley (1966) noted that "within the limits of confidence stated by the tests of significance, randomization can suffice without the pretest" (p. 25). Thus, according to Campbell and Stanley, it may not even be necessary to use matched groups because random assignment can be used to equate the groups.



What is the major exception to Campbell and Stanley's argument that randomization will create equal groups?

Random selection A control technique that ensures that each member of the population has an equal chance of being chosen for an experiment.

Selection A threat to internal validity that can occur if participants are chosen in such a way that the groups are not equal before the experiment; the researcher cannot then be certain that the IV caused any difference observed after the experiment.

History A threat to internal validity; refers to events that occur between the DV measurements in a repeated-measures design.

Maturation An internal validity threat; refers to changes in participants that occur over time during an experiment; could include actual physical maturation or tiredness, boredom, hunger, and so on.

Testing A threat to internal validity that occurs because measuring the DV causes a change in the DV.

Statistical regression This threat to internal validity occurs when low scorers improve or high scorers fall on a second administration of a test solely as a result of statistical reasons.

We hope that you remembered (from Chapters 10–12) that randomization is supposed to create equal groups *in the long run*. You should be aware of randomization's *possible* shortcoming if you conduct an experiment with small numbers of participants. Although randomization may create equal groups with few participants, we cannot be as confident about this possibility as when we use large groups.

Finally, you should remember from Chapter 4 that random assignment is *not* the same as **random selection**. Random assignment is related to the issue of internal validity; the notion of random selection is more involved with external validity (see Chapter 8).

Experimental Design Campbell and Stanley (1966) reviewed six experimental designs and evaluated them in terms of controlling for internal validity. They recommended three of the designs as being able to control the threats to internal validity we listed in Chapter 8. Let's examine their three recommended designs.

The Pretest–Posttest Control-Group Design The pretest–posttest control-group design appears in Figure 13-2. As you can see, this design consists of two randomly assigned groups of participants, both of which are pretested, with one group receiving the IV.

The threats to internal validity, which we summarized in Chapter 8, are controlled by one of two mechanisms in this design. The random assignment of participants to groups allows us to assume that the two groups are equated before the experiment, thus ruling out **selection** as a problem. Using a pretest and a posttest for *both* groups allows us to control the effects of **history**, **maturity**, and **testing** because they should affect both groups equally. If the control group shows a change between the pretests and posttests, then we know that some factor other than the IV is at work. **Statistical regression** is controlled as long as we assign our experimental and control groups from the same extreme pool of participants. If any of the **interactions with selection** occur, they should affect both groups equally, thus equalizing those effects on internal validity.

The other threats to internal validity are not controlled, but the pretest–posttest control-group design does give us the ability to determine whether they were

R	O ₁	O ₂	(control group)
R	O ₃	X	O ₄ (experimental group)
KEY:			
R = Random assignment			
O = Pretest or posttest observation or measurement			
X = Experimental variable or event			
Each row represents a different group of participants.			
Left-to-right dimension represents passage of time.			
Any letters vertical to each other occur simultaneously.			
(Note: This key also applies to Figures 13-3, 13-5, and 13-6)			

FIGURE 13-2 The Pretest–Posttest Control-Group Design.

problematic in a given experiment. We can check to see whether **experimental mortality** was a problem because we measure both groups on two occasions. **Instrumentation** is measured if we are dealing with responses to a test, for example, but it could still remain a problem if human interviewers or observers are used. There is simply no substitute for pretraining when you use humans to record or score data for you. Finally, **diffusion or imitation of treatments** could still remain a problem if participants from the control group (or different experimental groups) learn about the treatments for other groups. Again, though, you do have the control group as a “yardstick” to determine whether their scores increase or decrease in similar fashion to the experimental group’s scores. If you see similar changes, you can suspect that internal validity controls may have failed.

The Solomon Four-Group Design Figure 13-3 contains a diagram of the Solomon four-group design, first proposed by Solomon (1949). Notice that this design is identical to the pretest–posttest control-group design with the first two groups but adds an additional two groups, thus gaining the name *four-group design*. Because the Solomon four-group design has the same two groups as the pretest–posttest control-group design, it has the same protection against the threats to internal validity. The main advantage gained by adding the two additional groups relates to external validity (see Chapter 8).

One problem with the Solomon design comes in conducting statistical analysis of the data because there is no statistical test that can treat all six sets of data at the same time. Campbell and Stanley (1966) suggested treating the posttest scores as a factorial design, as shown in Figure 13-4. Unfortunately, this approach ignores all the pretest scores.

The Posttest-Only Control-Group Design Figure 13-5 shows the posttest-only control-group design. As you can see by comparing Figure 13-5 to Figures 13-2 and 13-3, the posttest-only control-group design is a copy of the pretest–posttest control-group design, without the pretests included, and is a copy of the two added groups in the Solomon four-group design. Does the lack of pretests render the posttest-only control-group design less desirable than the other two designs that include them? No, because we can count on the random assignment to groups to equate the two groups. Thus, using random assignment of participants to groups and withholding the IV from one group to make it a control group is a powerful experimental design that controls the threats to internal validity we covered in Chapter 8.

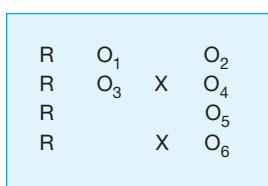


FIGURE 13-3 The Solomon Four-Group Design.
This design is used to protect internal validity.

Interactions with selection

These threats to internal validity can occur if there are systematic differences between or among selected treatment groups based on maturation, history, or instrumentation.

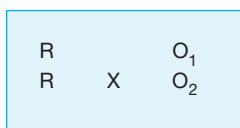
Experimental mortality

This threat to internal validity can occur if experimental participants from different groups drop out of the experiment at different rates.

Instrumentation This threat to internal validity occurs if the equipment or human measuring the DV changes its measuring criterion over time.

Diffusion or imitation of treatment This threat to internal validity can occur if participants in one treatment group become familiar with the treatment being received by another group and copy that treatment.

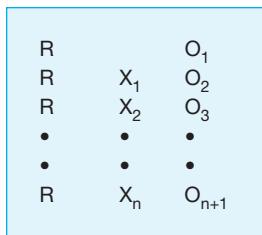
	No IV	Receives IV
Pretested	O ₂	O ₄
Unprettested	O ₅	O ₆

FIGURE 13-4 Factorial Treatment of Solomon Four-Group Design Posttest Scores.**FIGURE 13-5** Posttest-Only Control-Group Design.
This is a powerful design for protecting internal validity.

After examining Figure 13-5, what type of design (from Chapters 10–12) does this appear to be?

We hope that you identified Figure 13-5 as the two-group design from Chapter 10. We must point out, however, that it is *not* critical to have only two groups in this design. The posttest-only control-group design could be extended by adding additional treatment groups, as shown in Figure 13-6. This extended design should remind you of the multiple-group design discussed in Chapter 11.

Finally, we could create a factorial design from the posttest-only control group by combining two of these designs simultaneously so that we ended up with a block diagram similar to those from Chapter 12.

**FIGURE 13-6** An Extension of the Posttest-Only Control-Group Design. This design permits testing of multiple treatment groups.



It should be clear that the posttest-only design is not defined by the number of groups. What is (are) the defining feature(s) of this design? Take a moment to study Figures 13-5 and 13-6 before answering.

The two features that are necessary to “make” a posttest-only control-group design are random assignment of participants to groups and the inclusion of a control (no-treatment) group. These features allow the design to derive cause-and-effect statements by equating the groups before the experiment and controlling the threats to internal validity.

We hope you can appreciate the amount of control that can be gained by the two simple principles of random assignment and experimental design. Although these principles are simple, they are quite elegant in the power they bring to the experimental situation. You would be wise not to underestimate their importance.

Conclusion

How important is internal validity? It is *the most important* property of any experiment. If you do not concern yourself with the internal validity of your experiment, you are wasting your time. Experiments are intended to produce cause-and-effect statements—to conclude that X causes Y to occur. If you merely wish to learn something about the association of two variables, you can use one of the nonexperimental methods for acquiring data summarized in Chapter 4 or calculate a correlation coefficient. If you wish to investigate the cause(s) of a phenomenon, you must control any extraneous variables that might affect your dependent variable. You cannot count on your statistical tests to provide the necessary control functions for you. Statistical tests merely analyze the numbers you bring to the test; they do not have the ability to remove confounding effects (or even to discern that confounding has occurred) in your data.

■ REVIEW SUMMARY

1. One important control for **internal validity** is **random assignment** of participants to groups. This procedure assures us that the groups are equated before beginning the experiment.
2. **Random selection** refers to choosing our participants from a population so that all potential participants could be chosen. Random selection is important to external validity.
3. The pretest–posttest control-group design consists of two groups of participants that have been randomly assigned to an experimental and control group, pretested and posttested, with the experimental group receiving the IV. This design controls for internal validity threats but has the problem of including a pretest.
4. The Solomon four-group design is a copy of the pretest–posttest control-group design except that it adds two groups that have not been pretested. This design also controls for internal validity threats, but there is no statistical test that can be used to analyze all six sets of data.

5. The posttest-only control-group design consists of two groups of participants that are randomly assigned to experimental and control groups, with the experimental group receiving the IV treatment. Both groups are tested with a posttest. This design controls for internal validity threats and is free from other problems.
6. The posttest-only control-group design can be extended to include additional treatment groups or additional IVs.
7. It is essential for an experiment to be internally valid; otherwise, no conclusion can be drawn from the experiment.

Check Your Progress

1. The two general methods we use to protect the internal validity of our experiment are _____ and _____.
2. Why is it essential to use random assignment of our participants to their groups?
3. Distinguish between random assignment and random selection.
4. What is the drawback of using the pretest–posttest control-group design to help with internal validity?
5. A friend tells you she was a participant in a psychology experiment and says, “It was crazy! We took a personality test, watched a film, and then took the same test again!” From this description, you could tell that she was in the
 - a. control group of a posttest-only control-group design
 - b. experimental group of a posttest-only control-group design
 - c. control group of a pretest–posttest control-group design
 - d. experimental group of a pretest–posttest control-group design
6. What is the drawback of using the Solomon four-group design as a control for internal validity?
7. Diagram the posttest-only control-group design. Why is it a good choice for controlling internal validity?

Single-case experimental design An experiment that consists of one participant (also known as $N = 1$ designs).

Single-Case Experimental Designs

A **single-case experimental design** (also known as an $N = 1$ design) is just that. This term simply refers to an experimental design with one participant. This approach, of course, is quite similar to the detective’s strategy of pursuing a single suspect.



The $N = 1$ approach probably sounds familiar to you. What data-gathering approach have we studied that involves one participant?



Reprinted with special permission of King Features Syndicate and Tom Cheney.

"Sooner or later he'll learn that when he presses the bar, he'll receive a salary."

Much psychological knowledge has been gained from single-case designs.

We hope you remember the **case-study approach** from Chapter 3. In a case study we conduct an intense observation of a single individual and compile a record of those observations. As we noted in Chapter 3, case studies are often used in clinical settings. If you have taken an abnormal psychology course, you probably remember reading case studies of people with various disorders. The case study is an excellent descriptive technique; if you read a case study about an individual with a mental disorder, you get a vivid picture of that disorder. On the other hand, a case study is *merely* a descriptive or observational approach; the researcher does not manipulate or control variables but simply records observations. Thus, case studies do not allow us to draw cause-and-effect conclusions.

You will remember that we must institute control over the variables in an experiment in order to derive cause-and-effect statements. In a single-case design we institute controls just as we do in a typical experiment—the only difference is that our experiment deals with just one participant. Also, just as in a typical experiment, we must take precautions in dealing with the internal validity of a single-case design. We hope that the single-case design raises many questions for you. After all, it does go against the grain of some of the principles we have developed thus far. Let's take a quick look at this design's history and uses, which will help you understand its importance.

Case-study approach

An observational technique in which a record of observations about a single participant is compiled.

History of Single-Case Experimental Designs

The single-case experimental design has quite an illustrious past in experimental psychology (Hersen, 1982; Hersen & Barlow, 1976). In the 1860s Gustav Fechner explored sensory processes through the use of psychophysical methods. Fechner developed two concepts that you probably remember from your introductory psychology course: *sensory thresholds* and the *just noticeable difference (jnd)*. Fechner conducted his work on an in-depth basis with a series of individuals. Wilhelm Wundt (founder of the first psychology laboratory) conducted his pioneering work on introspection with highly trained individual participants. Herman Ebbinghaus conducted perhaps the most famous examples of single-case designs in our discipline. Ebbinghaus was the pioneering researcher in the field of verbal learning and memory. His research was unique—not because he used the single-case design, but because he was the single participant in those designs. According to Dukes (1965), Ebbinghaus learned about 2,000 lists of nonsense syllables in his research over many years. Dukes provided several other examples of famous single-case designs with which you are probably familiar, such as Cannon's study of stomach contractions and hunger, Watson and Rayner's study of Little Albert's learned fears, and several researchers' work with language learning in individual apes.

Other than the ape-language studies cited by Dukes (1965), all these single-case design examples date to the 1800s and early 1900s. Dukes found only 246 single-case examples in the literature between 1939 and 1963. Clearly, there are fewer examples of single-case designs than group designs in the literature.



Can you think of a reason why single-case designs may have been more popular in the past?

Hersen (1982) attributed the preference for group designs over single-case designs to statistical innovations made by Sir Ronald A. Fisher. Fisher was a pioneer of many statistical approaches and techniques. Most important for this discussion, in the 1920s he developed analysis of variance (ANOVA; Spatz, 2001), which we covered in detail in Chapters 11 and 12. Combined with Gosset's early 1900s development of a test based on the *t* distribution (see Chapter 10), Fisher's work gave researchers a set of inferential statistical methods with which to analyze sets of data and draw conclusions. You may have taken these tests for granted and assumed that they had been around forever, but that is not the case. As these methods became popular and accessible to more researchers, the use of single-case designs declined. In today's research world, statistical analyses of incredibly complex designs can be completed in minutes (or even seconds) on computers you can hold in your hand. The ease of these calculations has probably contributed to the popularity of group designs over single-case designs.

Experimental analysis of behavior A research approach popularized by B. F. Skinner, in which a single participant is studied.

Uses of Single-Case Experimental Designs

There are still some researchers who use single-case designs. Founded by B. F. Skinner, the **experimental analysis of behavior** approach continues to employ this technique. Skinner (1966) summarized his philosophy in this manner: "Instead

of studying a thousand rats for one hour each, or a hundred rats for ten hours each, the investigator is likely to study one rat for a thousand hours" (p. 21). The Society for the Experimental Analysis of Behavior was formed and began publishing its own journals, the *Journal of the Experimental Analysis of Behavior* (in 1958) and the *Journal of Applied Behavior Analysis* (in 1968). Single-case designs are thus still used today; however, the number of users is small compared to those who use group designs, as you could guess by the handful of journal titles devoted to this approach.

One question that might occur to you is "Why use a single-case design in the first place?" Sherlock Holmes knew that "the world is full of obvious things which nobody by any chance ever observes" (Doyle, 1927, p. 745). Dukes (1965) provided a number of convincing arguments for and situations that require single-case designs. Let's look at several. First, a sample of one is all you can manage if that sample exhausts the population. If you have access to a participant who is unique, you simply cannot find other participants. Of course, this example is perhaps closer to a case study than to an experiment because there would be no larger population to which you could generalize your findings. Second, if you can assume perfect generalizability, then a sample of one is appropriate. If there is only inconsequential variability among members of the population on a particular variable, then measuring one participant should be sufficient. Third, a single-case design would be most appropriate when a single *negative* instance would refute a theory or an assumed universal relation. If the scientific community believes that "reinforcement always increases responding," then finding one instance in which reinforcement does *not* increase responding invalidates the thesis. Fourth, you may simply have limitations on your opportunity to observe a particular behavior. Behaviors in the real world (i.e., nonlaboratory behaviors) may be so rare that you can locate only one participant who exhibits the behavior. Dukes used examples of people who feel no pain, who are totally color-blind, or who exhibit dissociative identity disorder (again, close to a case study). You may remember reading about H. M. when you studied memory in introductory psychology. Because of the surgery for epilepsy that removed part of his brain, H. M. could no longer form new long-term memories. Researchers have studied H. M. for almost 50 years for clues about how the brain forms new memories (Corkin, 1984; Hilts, 1995). H.M.'s case was famous enough that the *New York Times* carried his obituary when he died in late 2008 (http://www.nytimes.com/2008/12/05/us/05hm.html?_r=2). Fifth, when research is extremely time consuming and expensive, requires extensive training, or presents difficulties with control, an investigator may choose to study just one participant. The studies in which researchers have attempted to teach apes to communicate through sign language, plastic symbols, or computers fall into this category. Obviously, there are instances in which a single-case design is totally appropriate.

General Procedures of Single-Case Experimental Designs

Hersen (1982) listed three procedures that are characteristic of single-case designs: repeated measures, baseline measurement, and changing one variable at a time. Let's see why each of these procedures is important.

Repeated Measures When we deal with many participants, we often measure them only once and then average all our observations. When you are dealing with only one participant, however, it is important to make sure that the behavior you are measuring is consistent. You would therefore repeatedly measure the participant's behavior. Control during the measurement process is extremely important. Hersen and Barlow (1976) noted that the procedures

for measurement "must be clearly specified, observable, public, and replicable in all respects" (p. 71). In addition, these repeated measurements "must be done under exacting and totally standardized conditions with respect to measurement devices used, personnel involved, time or times of day measurements are recorded, instructions to the subject, and the specific environmental conditions" (p. 71). Thus, conducting a single-case experiment and making repeated measurements do *not* remove the experimenter's need to control factors as carefully as possible.

Baseline A measurement of a behavior made under normal conditions (i.e., no IV is present); a control condition.

Baseline Measurement In most single-case designs the initial experimental period is devoted to determining the **baseline** level of behavior. In essence, baseline measurement serves as the control condition against which to compare the behavior as affected by the IV. When you are collecting baseline data, you hope to find a stable pattern of behavior so that you can more easily observe any change that occurs in the behavior after your intervention (IV). Barlow and Hersen (1973) recommended that you collect *at least* three observations during the baseline

period in order to establish a trend in the data. Although you may not achieve a stable measurement, the more observations you have, the more confident you can be that you have determined the general trend of the observations. Figure 13-7 depicts a hypothetical stable baseline presented by Hersen and Barlow (1976). Notice that they increased their odds of finding a stable pattern by collecting data three times per day and averaging those data for the daily entry.

Changing One Variable at a Time In a single-case design it is vital that, as the experimenter, you change only one variable at a time when you move from one phase of the experiment to the next.

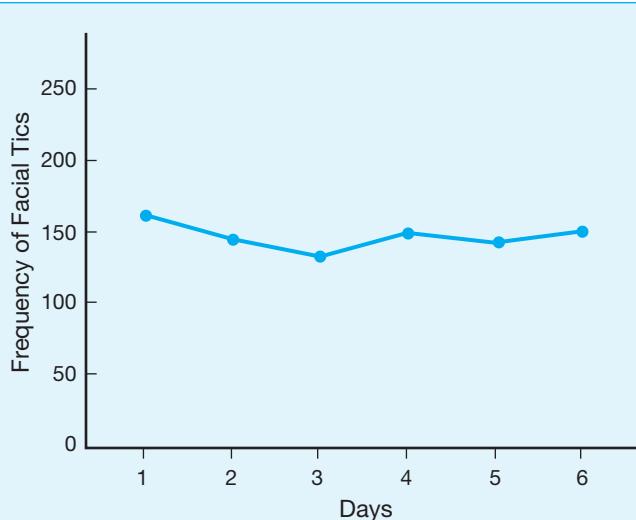


FIGURE 13-7 A Stable Baseline. Hypothetical data for mean number of facial tics averaged over three daily 15-minute videotaped sessions.

Source: Figure 3-1 from *Single-Case Experimental Designs: Strategies for Studying Behavioral Change*, by M. Hersen and D. H. Barlow, 1976, New York: Pergamon Press, p. 77. Used with permission of the publisher.



Why would it be important to change only one variable at a time in a single-case design?

We hope that the answer to this question came easily. Changing one variable at a time is a basic experimental control procedure that we have stressed many times. If you allow two variables to change simultaneously, then you have a confounded experiment and cannot tell which variable has caused the change in the behavior that you observe. This situation is exactly the same in a single-case design. If you record your baseline measurement, change several aspects of the participant's environment, and then observe the behavior again, you have no way of knowing which changed aspect affected the behavior.

Statistics and Single-Case Experimental Designs

Traditionally, researchers have not computed statistical analyses of results from single-case designs. Not only has the development of statistical tests for such designs lagged behind multiple-case analyses, but also there is controversy about whether statistical analyses of single-case designs are even appropriate (Kazdin, 1976). Both Kazdin (1976) and Hersen (1982) summarized the arguments concerning statistical analyses. Let's take a quick look at this controversy.

The Case Against Statistical Analysis As we mentioned, tradition and history say that statistical analyses are not necessary in single-case designs. The tradition has been to inspect visually ("eyeball") the data to determine whether change has taken place. Researchers who hold this position believe that treatments that do not produce visually apparent effects are either weak or ineffective. Skinner (1966) wrote that "rate of responding and changes in rate can be directly observed . . . [and] statistical methods are unnecessary" (p. 20).

Because many single-case studies involve clinical treatments, another argument against statistical analysis is that statistical significance is not always the same as clinical significance. A statistical demonstration of change may not be satisfying for practical application. "For example, an autistic child may hit himself in the head 100 times an hour. Treatment may reduce this to 50 times per hour. Even though change has been achieved, a much larger change is needed to eliminate behavior" (Kazdin, 1984, p. 89).

Finally, to the pro-statistics folks who argue that statistical analyses may help find effects that visual inspection would not (see next section), the anti-statistics camp makes the point that such subtle effects may not be replicable (Kazdin, 1976). As you saw in Chapter 8, if you cannot replicate a result, it has no external validity.

The Case for Statistical Analysis The argument for using statistical analyses of single-case designs revolves primarily around increased accuracy of conclusions. Jones, Vaught, and Weinrott (1977) have provided the most persuasive appeal for such analyses. They reviewed a number of studies published in the *Journal of Applied Behavior Analysis* that used visual inspection of data to draw conclusions. Jones et al. found that analyses of these data showed that sometimes conclusions drawn from visual inspections were correct and that sometimes

the conclusions were incorrect. In the latter category both Type I and Type II errors (see Chapter 9) occurred. In other words, some statistical analyses showed *no* effect when the researchers had said there was an effect, and some analyses showed significant effects when the researchers had said there were none. Kazdin (1976) pointed out that statistical analyses are particularly likely to uncover findings that do not show up in visual inspection when a stable baseline is not established, new areas of research are being investigated, or testing is done in the real world, which tends to increase extraneous variation.

As you can tell, there is no clear-cut answer concerning the use of statistics with single-case designs. Most researchers probably make their decision in such a situation based on a combination of personal preference, the audience for the information, and potential journal editors. Covering the various tests used to analyze single-case designs is beyond the scope of this text. Adaptations of *t* tests and ANOVA have been used, but these approaches have suffered from some problems. For further information about such tests, see Kazdin (1976).

Representative Single-Case Experimental Designs

A Refers to the baseline measurement in a single-case design.

B Refers to the outcome (treatment) measurement in a single-case design.

A-B design A single-case design in which you measure the baseline behavior, institute a treatment, and use a posttest.

Researchers use a standard notation for single-case designs that makes the information easier to present and conceptualize. In this notation, **A** refers to the baseline measurement and **B** refers to the measurement during or after treatment. We read the notation for single-case designs from left to right, to denote the passage of time.

A-B Design In the **A-B design**, the simplest of the single-case designs, we make baseline measurements, apply a treatment, and then take a second set of measurements. We compare the B (treatment) measurements to the A (baseline) measurements in order to determine whether a change has occurred. This design should remind you of a pretest–posttest design except for the absence of a control group. In the A-B design, the participant's A measurements serve as the control for the B measurements.

For example, Hall et al. (1971) used this approach in a special-education setting. A 10-year-old boy (Johnny) continually talked out and disrupted the class, which led other children to imitate him. The researchers asked the teacher to measure Johnny's baseline talking-out behavior (A) for five 15-minute sessions under normal conditions. In implementing the treatment (B), the teacher ignored the talking out and paid more attention to Johnny's productive behavior (attention was *contingent* on the desired behavior), again for five 15-minute sessions. Johnny's talking out diminished noticeably.

Hersen (1982) rated the A-B design as one of the weakest for inferring causality and noted that it is often deemed correlational.



Why do you think the A-B design is weak concerning causality?

The A-B design is poor for determining causality because of many of the threats to internal validity that we saw in Chapter 8. It is possible that another factor could vary along with the treatment. This possibility is especially strong for any extraneous variables that could be linked to time passage, such as history, maturation, and instrumentation. If such a factor varied across time with the treatment, then any change in B could be due to *either* the treatment or the extraneous factor. Because there is no control group, we cannot rule out the extraneous variable as a causative factor.



Can you think of a solution to the causality problem inherent in the A-B design? Remember that you cannot add a control group *or* participants because this is a single-case design. Any control must occur with the single participant.

The solution to this causality problem requires us to examine our next single-case design.

A-B-A Design In the **A-B-A design**, the treatment phase is followed by a return to the baseline condition. If a change in behavior during B is actually due to the experimental treatment, the change should disappear when B is removed and you return to the baseline condition. If, on the other hand, a change in B was due to some extraneous variable, the change will not disappear when B is removed. Thus, the A-B-A design allows a causal relation to be drawn.

In Hall et al.'s (1971) experiment, the teacher did return to the baseline condition with Johnny. When the teacher began again to pay attention to Johnny's talking-out behavior, that behavior increased considerably. This return to the previous behavior strengthened the researchers' claim that the treatment had caused the original decrease in Johnny's talking out.

A-B-A design A single-case design consisting of a baseline measurement, a treatment, a posttest, and a return to the baseline condition. It may not be recommended if the participant is left without a beneficial or necessary treatment in the second baseline.



There is one glaring drawback to the A-B-A design. Think about the implications of conducting a baseline–treatment–baseline experiment. Can you spot the drawback? How would you remedy this problem?

If you end your experiment on an A phase, this leaves the participant in a baseline condition. If the treatment is a beneficial one, the participant is "left hanging" without the treatment. The solution to this problem requires us to examine another single-case design.

On the other hand, returning to an A phase can give the researcher an idea of how effective a treatment was. Aurelie Welterlin (2004), a student at the University of North Carolina Chapel Hill, worked with a 7-year-old boy diagnosed with autism who exhibited impaired social interaction skills. Welterlin had the boy and two female peers play in a room together and measured the number of times the boy interacted with the peers. During the baseline

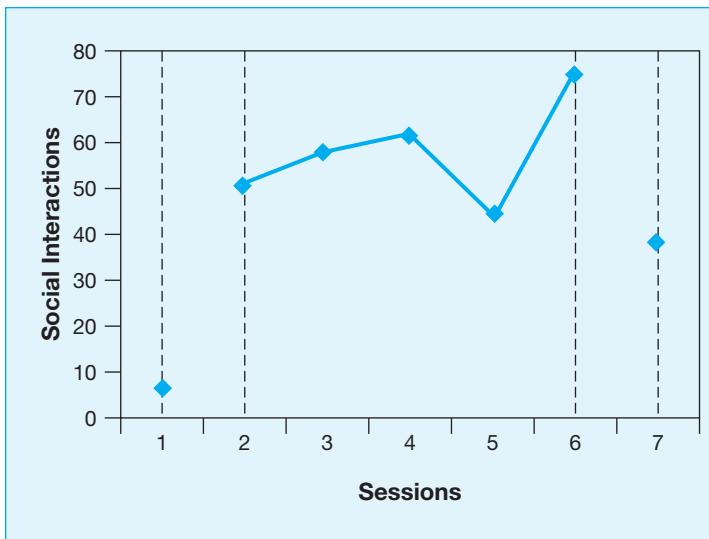


FIGURE 13-8 A-B-A Design. Teaching a 7-Year-Old Boy to Increase Social Interaction. Sessions 1 and 7 represent baseline periods; Sessions 2–6 used cuing to prompt interaction.

Source: Adapted from “Social Initiation in Children with Autism: A Peer Playgroup Intervention,” by A. Welterlin, 2004, *Psi Chi Journal of Undergraduate Research*, 9, pp. 97–104. Copyright © 2004 Psi Chi, The National Honor Society in Psychology (www.psichi.org). Reprinted by permission. All rights reserved.

period (A; Session 1), it was highly unusual for the boy to interact at all (see Figure 13-8). During the intervention sessions (B; Sessions 2–6), a facilitator cued the boy in an attempt to get him to interact with the girls. During the second baseline (A; Session 7), the facilitator did not provide any cues to the boy. An examination of Figure 13-8 shows that the intervention (cuing) did produce more social interaction from the boy. The second baseline period shows that, without the cues, the boy’s social interaction decreased; however, it remained higher than it had been in the original baseline. Welterlin was thus able to demonstrate that cuing did increase social interaction (B) and that its effects persisted even in the absence of the cuing (second baseline).

A-B-A-B design A single-case design consisting of a baseline, treatment, posttest, return to baseline, repeated treatment, and second posttest. This design gives the best chance of isolating causation.

A-B-A-B Design As you can figure out by now, the **A-B-A-B design** begins with a baseline period followed by treatment, baseline, and treatment periods consecutively. This design adds a final treatment period to the A-B-A design, thereby completing the experimental cycle with the participant in a treatment phase. Hersen and Barlow (1976) pointed out that this design gives two transitions (B to A and A to B), which can demonstrate the effect of the treatment variable. Thus, our ability to draw a cause-and-effect conclusion is further strengthened.

Hall et al. (1971) actually used the A-B-A-B design in their experiment with Johnny. After measuring Johnny’s baseline talking-out behavior (A) under normal conditions, the teacher implemented the treatment (B) by ignoring the talking out and paying attention only to Johnny’s productive behavior. The teacher then repeated the A and B phases. Results from this study appear in Figure 13-9. This graph shows us several things. First, visual inspection of these results should be enough to convince us of the efficacy of the treatment—the difference between baseline and treatment conditions is dramatic. This graph is a good

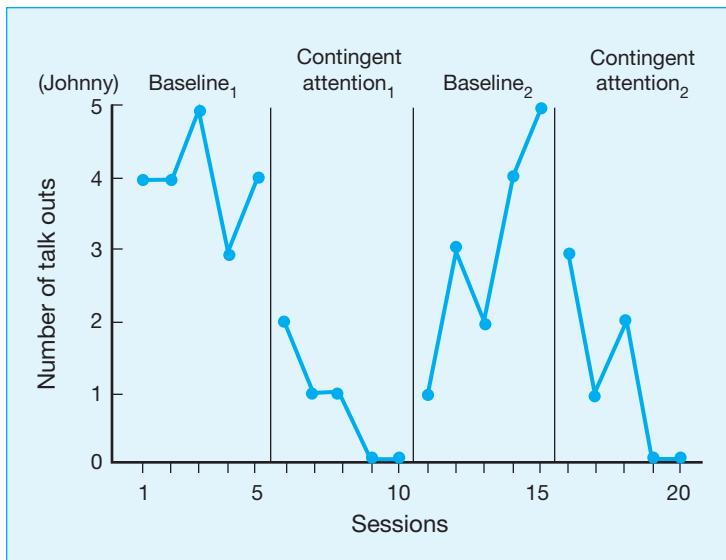


FIGURE 13-9 Talking-Out Behavior in a Mentally Retarded Student. A record of talking-out behavior of an educable mentally retarded student. **Baseline₁**: before experimental conditions; **Contingent Attention₁**: systematic ignoring of talking out and increased teacher attention to appropriate behavior; **Baseline₂**: reinstatement of teacher attention to talking-out behavior; **Contingent Attention₂**: return to systematic ignoring of talking out and increased attention to appropriate behavior.

Source: Figure 2 from "The Teacher as Observer and Experimenter in the Modification of Disrupting and Talking-out Behaviors," by R. V. Hall, R. Fox, D. Willard, L. Goldsmith, M. Emerson, M. Owen, et al., 1971, *Journal of Applied Behavior Analysis*, 4, p. 143.

illustration of why many researchers who use single-case designs believe that statistics are unnecessary. Second, it is apparent that the treatment did work. When the teacher stopped attending to Johnny's talking-out behavior and paid attention to his productive behavior, the talking out decreased substantially. Third, we can determine that the increased productive behavior was caused by the contingent attention because of the rapid increase in talking out when the attention was removed (see Baseline₂ in Figure 13-9).

Design and the Real World From the preceding sections it should be clear that the A-B-A-B design is the preferred design for single-case research; however, we must ask whether typical practice actually follows the recommended path. Hersen and Barlow (1976) acknowledged that researchers often use the A-B design despite its shortcomings in terms of demonstrating causality. The main reason the A-B design is used concerns either the inability or undesirability to return to the baseline in the third stage. In the real world, perfect experimental design cannot always be used. We must simply accept that our ability to draw definitive conclusions in such instances is limited. Let's look at three common situations that preclude using a design other than the A-B design.

First, as is typical in many field experiments, it may be impractical to reverse a treatment. Campbell (1969, p. 410) urged politicians to conduct social reforms as experiments, proposing that they initiate a new policy on an experimental basis. If after five years there had been no significant improvement, he recommended that the politicians shift to a different policy.

Political realities, of course, would not allow social change to be conducted experimentally. Campbell provided a good example of this problem. In 1955 Connecticut experienced a record number of traffic fatalities. The governor instituted a speeding crackdown in 1956, and traffic fatalities fell by more than 12%. Once this result occurred, it would have been politically stupid for the governor to announce, "We wish to determine whether the speeding crackdown actually caused the drop in auto deaths. Therefore, in 1957 we will relax our enforcement of speeding laws to find out whether fatalities increase once again." Yet this change is what would be necessary in order to rule out rival hypotheses and draw a definitive cause-and-effect statement.

Second, it may be unethical to reverse a treatment. Lang and Melamed (1969) worked with a 9-month-old boy (see Figure 13-10) who had begun vomiting after meals when he was about 6 months old. Doctors had implemented dietary changes, conducted medical tests, performed exploratory surgery, but could find no organic cause. The boy weighed 9 pounds, 4 ounces at birth, grew to 17 pounds at 6 months of age, but weighed only 12 pounds at 9 months. The child was being fed through a nose tube and was in critical condition (see Figure 13-10A). Lang and Melamed instituted a treatment consisting of brief and repeated shocks applied to the boy's leg at the first signs of vomiting and ending when vomiting ceased. By the third treatment session, one or two brief shocks were enough to stop the vomiting. By the fourth day of

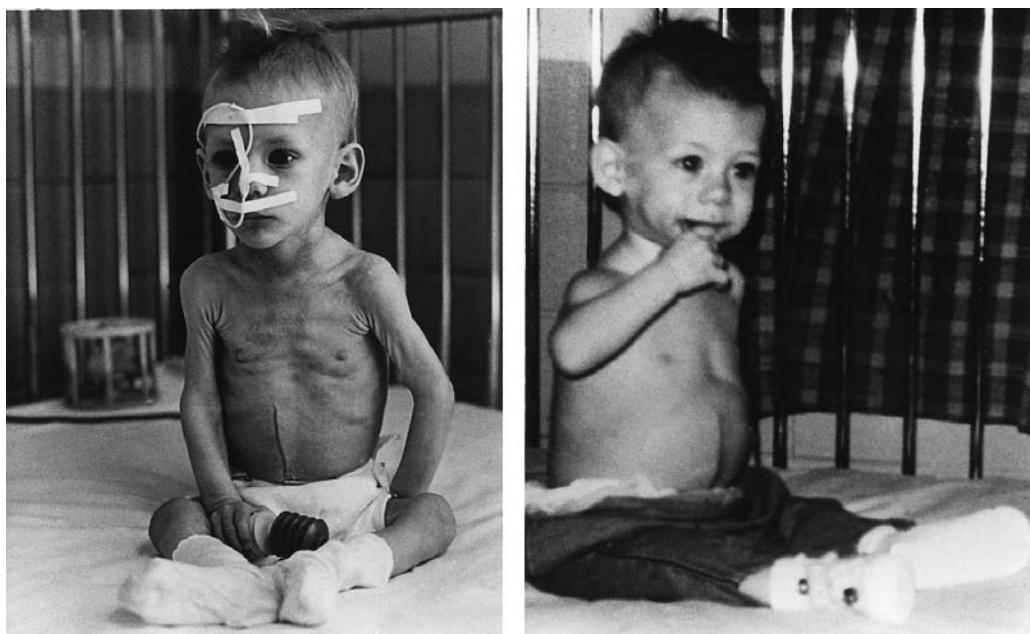


FIGURE 13-10 Nine-Month-Old Boy Hospitalized for Frequent Vomiting (**A**) Before Treatment and (**B**) After Treatment (13 Days Later). The photograph at the left was taken during the observation period just prior to treatment. (It clearly illustrates the patient's debilitated condition: lack of body fat, skin hanging in loose folds. The tape around the face holds tubing for the nasogastric pump. The photograph at the right was taken on the day of discharge from the hospital, 13 days after the first photo. The 26% increase in body weight already attained is easily seen in the full, more infantlike face, the rounded arms, and more substantial trunk.)

Source: Figure 1 from "Avoidance Conditioning Therapy of an Infant with Chronic Ruminative Vomiting," by P. J. Lang and B. G. Melamed, 1969, *Journal of Abnormal Psychology*, 74, pp. 1-8.

treatment, vomiting stopped and treatment was discontinued. Two days later, some vomiting occurred, so the procedure was reinstated for three sessions. Five days later, the child was dismissed from the hospital (see Figure 13-10B). A month later, he weighed 21 pounds, and 5 months later weighed over 26 pounds, with no recurrence of vomiting. Although this treatment bears some resemblance to an A-B-A-B design (because of the brief relapse), the additional session was not originally intended and was *not* conducted as an intentional removal of B to chart a new baseline—the researchers believed that the problem had been cured at the point treatment was discontinued. We are certain that you can see why ethical considerations would dictate an A-B design in this instance rather than the more experimentally rigorous A-B-A-B design.

Finally, it may be impossible, undesirable, or unethical to reverse a treatment if learning takes place during the treatment. Bobby Traffanstedt (1998), a student at the University of Central Arkansas in Conway, used an A-B design to modify a 10-year-old boy's TV watching and exercise behaviors. Traffanstedt wanted to teach the boy to spend less time watching TV and more time exercising. He used the operant procedures of shaping and reinforcement while working with the child for several weeks. The baseline (Week 1) and posttest (Weeks 2–9) behavior measures appear in Figure 13-11. As you can see, visual inspection of these data is convincing.

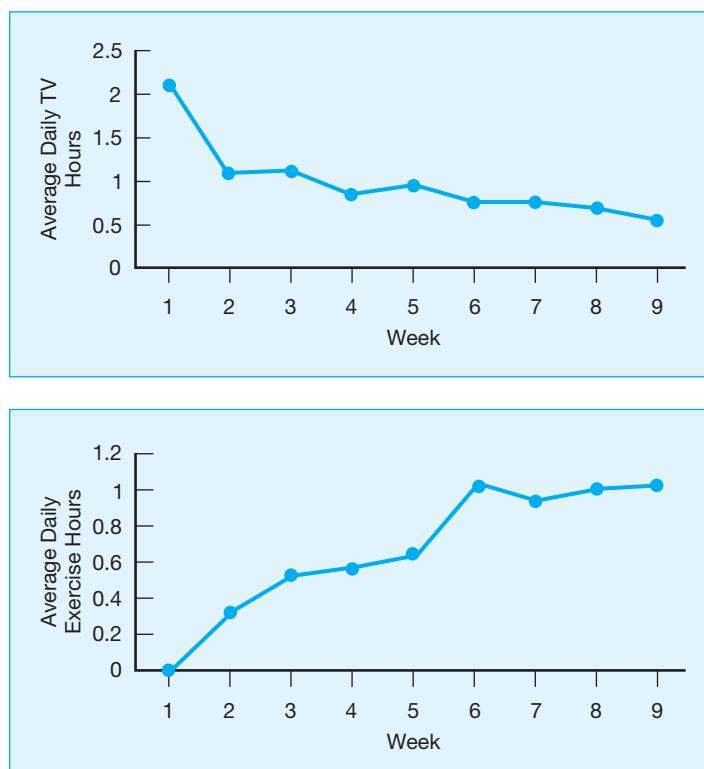


FIGURE 13-11 Teaching a 10-Year-Old Boy to Decrease TV Viewing and Increase Exercise. Traffanstedt (1998) used shaping and reinforcement to modify the boy's behavior.

Source: Adapted from "Weight Reduction Using Behavior Modification" by B. Traffanstedt, 1998, *Journal of Psychological Inquiry*, 3, pp. 19–23.



Why did Traffanstedt (1988) *not* use the A-B-A-B design in this experiment?

Because Traffanstedt had successfully taught the child to spend less time watching TV and more time exercising, he did not want to “undo” this learning and return to the baseline condition. Traffanstedt had already attempted to fade out the reinforcement gradually over the course of Weeks 2 through 9; going back to baseline was not really feasible. Having learned the new behaviors, it would make no sense to return the boy to the baseline condition.

The conclusion to this section is that you as an experimenter may find yourself caught in the middle. On the one hand, you have the knowledge of proper experimental design and what is necessary to yield cause-and-effect explanations. On the other hand, you have the realities of applied situations. The best rule of thumb for such situations is that you should use the most stringent experimental design you can, but you should not give up on an important project if you cannot use the absolute best design that exists. As a psychological detective, you have an edge on the real-life detective, who cannot apply a design even as rigorous as those we have presented in this section. The police detective must always work on a solution after the fact.

Additional Single-Case Designs In presenting the A-B, A-B-A, and A-B-A-B designs, we have merely scratched the surface of single-case designs. We have covered the designs we think you might be likely to use in the near future. As our references show, entire books have been written about single-case designs. Hersen and Barlow (1976) covered many additional variations on single-case designs, including designs with multiple baselines, multiple schedules, and interactions. Thus, if you ever envision a single-case design that is more complicated than the ones we have presented in this text, we refer you to Hersen and Barlow or a similar book dealing with single-case designs.

■ REVIEW SUMMARY

1. **Single-case experimental designs** are experiments that deal with a single participant.
2. Single-case designs have several legitimate uses.
3. Single-case designs are characterized by repeated measures, **baseline** measurement, and changing one variable at a time.
4. There is controversy over the use of statistics with single-case designs. The traditional approach has been to draw conclusions by visually examining the data. Proponents of statistical analysis maintain that analysis yields more accurate conclusions.
5. The **A-B-A-B** single-case design allows you the best chance to draw a cause-and-effect conclusion regarding a treatment. Realities of the real world often force the use of **A-B** designs, which are particularly prone to alternative explanations.

Check Your Progress

1. Why were single-case designs quite popular in psychology's early years but less popular today?
2. How can a single-case design be used to disprove a theory?
3. To come up with a comparison in the single-case design, we first measure behavior before the treatment during the _____ period. To get a stable measurement, we should make at least _____ observations.
4. In essence, _____ serve(s) as the control condition in the single-case design.
 - a. baseline measurements
 - b. repeated measures
 - c. changing one variable at a time
 - d. experimental analysis of behavior
5. Summarize two arguments for and two arguments against the use of statistical analysis in single-case designs.
6. Match the design with the appropriate characteristic.

1. A-B	A. leaves the participant in a baseline phase
2. A-B-A	B. best single-case design for determining cause-and-effect relations
3. A-B-A-B	C. has many threats to internal validity
7. Why might you be forced to use an A-B single-case design in the real world? Give an original example of such a situation.

Quasi-Experimental Designs

In this section we will deal with designs that are virtually identical to true experimental designs with the *exception* of random assignment of participants to groups. When we are able to manipulate an IV and measure a DV but *cannot* randomly assign our participants to groups, we must use a **quasi-experimental design**. Similarly, police detectives sometimes face situations in which they must build their case on circumstantial evidence rather than on direct evidence.

Quasi-experimental design A research design used when the researcher cannot randomly assign experimental participants to the groups but the researcher does manipulate an IV and measure a DV.



What problem results when we cannot randomly assign research participants to groups?

Not being able to assign our participants randomly to their groups has the effect of violating an important assumption that allows us to draw cause-and-effect conclusions from our experiments—the assumption of equal groups before the experiment. Even if we can randomly select participants from a larger group, we cannot make cause-and-effect statements

without random assignment. For example, you could randomly *select* students from an introductory psychology course, but you could not randomly *assign* them to groups based on sex! As Campbell and Stanley (1966) pointed out, the assumption of random assignment has been an important part of statistics and experimental design since the time of Fisher. If we unknowingly began an experiment with unequal groups and our statistics showed a difference after the experiment, we would make a Type I error (see Chapter 9) by concluding that the IV caused the difference that was actually present from the outset. Clearly, this conclusion could be wrong.

It is likely that our description of quasi-experimental design reminds you of the ex post facto studies we covered in Chapter 4. Some writers categorize ex post facto and quasi-experimental designs together and some separate them. We will draw a small, but significant, distinction between the two. Remember, in Chapter 4 we described the ex post facto study as having an IV that had *already* occurred and could not be manipulated. Thus, if we wish to study sex differences on mathematics or English achievement, we are studying the IV of biological sex, which we cannot control or manipulate. Of course, because the IV is a pre-existing condition, we also cannot randomly assign our participants to groups.

On the other hand, in a quasi-experimental design our participants belong to preexisting groups that cannot be randomly assigned; however, we *do* have control over the IV—we can administer it when and to whom we wish. Thus, we could choose our participants on the basis of sex and *then* have some of them participate in a workshop designed to improve their math or English achievement. In this case the workshop (or lack thereof) would serve as the IV for the preexisting groups of boys and girls, and the math or English achievement scores would be the DV. Obviously, random assignment is impossible in this case. Quasi-experimental designs are a step closer to true experimental designs than ex post facto studies because you, as the experimenter, are able to exert control over the IV and its administration. Being able to administer your own IV is preferable to having nature administer it for you, at least in terms of control.

The basic rationale for using quasi-experimental designs is the same as that for ex post facto studies—your inability to assign participants at random. According to Hedrick, Bickman, and Rog (1993), “[A] quasi-experimental design is not the method of choice, but rather a fall-back strategy for situations in which random assignment is not possible” (p. 62). When dealing with selection variables that do not allow for random assignment, we have the choice of using a quasi-experimental design or simply ignoring an important or interesting experimental question. Instead of letting such questions go unasked, researchers resort to quasi-experimental research.

History of Quasi-Experimental Designs

It is difficult to trace the history of quasi-experimental designs. Although McGuigan (1960) did not include the term in the first edition of his classic experimental psychology text, Campbell and Stanley did use it in the title of their 1966 guide to experimental design. There is little doubt, however, that researchers were tackling quasi-experimental design problems long before Campbell and Stanley's published work. Cook and Campbell (1979) noted that some researchers were writing about quasi-experiments in the 1950s, although the term did not originate until later. It is likely that Campbell and Stanley (1966) and Cook and Campbell (1979) are responsible for elevating quasi-experimental work to the respectable position it holds today.

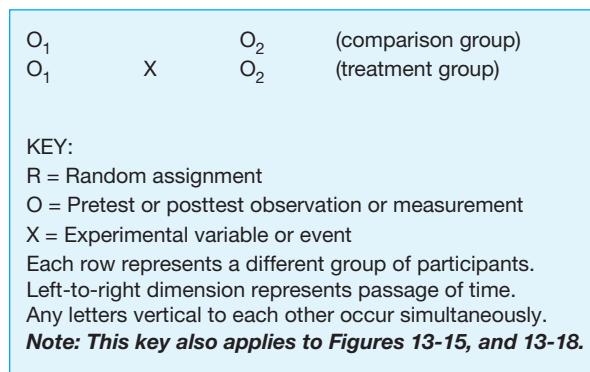
Uses of Quasi-Experimental Designs

Hedrick et al. (1993) listed several specific situations that require quasi-experimental designs. Let's take a brief look at their list. First, there are many variables that simply make random assignment impossible. If we wish to study participants from certain groups (e.g., based on sex, age, previous life experiences, personality characteristics), we must use quasi-experimental designs. Second, when you wish to evaluate an ongoing program or intervention (a retrospective study), you would have to use a quasi-experimental design. Because the program began before you decided to evaluate it, you would have been unable to use control procedures from the outset. Third, studies of social conditions demand quasi-experimental designs. You would not study the effects of poverty, race, unemployment, or other such social factors through random assignment. Fourth, it is sometimes the case that random assignment is not possible because of expense, time, or monitoring difficulties. For example, if you conducted a cross-cultural research project involving participants from several different countries, it would be nearly impossible to guarantee that the same random assignment procedures were used in each setting. Fifth, the ethics of an experimental situation, particularly with psychological research, may necessitate quasi-experimentation. For example, if you are conducting a research program to evaluate a certain treatment, you must worry about the ethics of withholding that treatment from people who could benefit from it. As you will see, quasi-experimentation provides a design that will work in such situations to remove this ethical dilemma.

Representative Quasi-Experimental Designs

Unlike the single-case design, we do not include sections covering general procedures and statistics of quasi-experimental designs. It is difficult to derive general principles because the representative designs we are about to introduce are so varied in nature. Because quasi-experimental designs resemble true experiments, the use of statistics for quasi-experimental designs is not an issue; the traditional statistical tests used with true experiments are also appropriate for quasi-experiments.

Nonequivalent Group Design The **nonequivalent group design** (Campbell & Stanley, 1966) appears in Figure 13-12.



Nonequivalent group design A design involving two or more groups that are not randomly assigned; a comparison group (no treatment) is compared to one or more treatment groups.

FIGURE 13-12 The Nonequivalent Group Design.



The nonequivalent group design should remind you of a design that we covered in the section on research designs that protect internal validity. Which design does it resemble? How is it different? What is the implication of this difference?

If you turn back to Figure 13-2, you will see that the nonequivalent group design bears a distinct resemblance to the pretest–posttest control-group design; however, the nonequivalent group design is missing the *R*s in front of the two groups; random assignment is *not* used in creating the groups. The lack of random assignment means that our groups may differ before the experiment—thus the name *nonequivalent group* design.

You also will notice that the two groups are labeled as the *comparison group* (rather than control group) and the *treatment group* (rather than experimental group [from Hedrick et al., 1993]). The *treatment* to *experimental* change is not particularly important; those terms could be used interchangeably; however, changing the name from *control* to *comparison* group is important and meaningful. In the nonequivalent group design, this group serves as the comparison to the treatment group but cannot truly be called a control group because of the lack of random assignment.

It is possible to extend the nonequivalent group design to include more than one treatment group if you wish to contrast two or more treatment groups with your comparison group. The key to the nonequivalent group design is creating a good comparison group. As far as is possible, we attempt to create an equal group through our selection criteria rather than through random assignment.

Examples of procedures for creating such a group include using members of a waiting list for a program/service; using people who did not volunteer for a program, but were eligible; using students in classes that will receive the curriculum (treatment) at a later date; and matching individual characteristics. (Hedrick et al., 1993, p. 59)

Geronimus (1991) provided a good example of creating a strong comparison group. She and her colleagues completed several studies of long-term outcomes for teen mothers. As you are probably aware, the stereotypical outcome for teen mothers is quite dismal: Younger mothers are more likely to have negative experiences such as poverty, high dropout rates, and higher rates of infant mortality. Geronimus believed that family factors, such as socioeconomic status, might be better predictors of these negative outcomes than the actual teen pregnancy. Random assignment for research on this topic would be impossible—you could not randomly assign teenage girls to become pregnant. Quasi-experimentation was thus necessary. In looking for a comparison group that would be as similar as possible, Geronimus decided to use the teenage mothers' sisters who did not become pregnant until later in life. Thus, although the assignment to groups was not random, the groups were presumably very near to equivalence, particularly with respect to family background factors. Interestingly enough, when family background was controlled in this manner, many of the negative outcomes associated with teen pregnancy disappeared. For example, there was no longer any difference in the dropout rates of the two groups. “For indicators of infant health and children’s sociocognitive development, at times the trends reversed direction (i.e., controlling for family background, the teen birth group did better than the postponers)” (Geronimus, 1991, p. 465).

In Geronimus's research the "pretest" (actually a matching variable in this case) consisted of finding two women from the same family, one who first became pregnant as a teenager and one who did not get pregnant until after age 20. In this case the groups may still have been nonequivalent, but they were highly equivalent on family background. Sometimes it is impossible to begin with equivalent groups, and the pretest serves much like a baseline measure for comparison with the posttest. In this type of situation the label *nonequivalent groups* seems quite appropriate.

Janet Luehring, a student at Washburn University in Topeka, Kansas, and Joanne Altman, her faculty advisor, used a nonequivalent group design in their research project (Luehring & Altman, 2000). They measured students' performance on the Mental Rotation Task (MRT; Vandenberg & Kuse, 1978). For each item on the MRT, participants saw five three-dimensional shapes, with the first shape being the test stimulus. Two of the other four shapes were matches of the test stimulus when rotated; participants had to identify the two that were the same as the test stimulus. The MRT consists of 20 such items and normally has a 6-minute time limit. The preponderance of evidence from psychological research indicates that men tend to perform better on spatial tasks than women (Luehring & Altman, 2000). Luehring and Altman compared the performance of female students on the MRT to that of male students; the groups, thus, were not equal before the experiment began. The IV in Luehring and Altman's experiment consisted of performing the MRT under timed or untimed conditions. They found that women who performed the MRT under timed conditions made as few errors as men under timed or untimed conditions—only the women under untimed conditions made more errors than the other three groups. Because the two gender groups began the experiment as nonequivalent, the appropriate question after the experiment was not whether a difference existed, but whether the difference was the same as before the experiment (see Figure 13-13A) or whether the difference had changed in some way (see Figure 13-13B). In Luehring and Altman's experiment the difference between the two groups had grown smaller in the timed condition, thus supporting the hypothesis that the IV had an effect on MRT performance for women. Of course, there are several other possible outcomes that would show some effect of the IV. More of Cook and Campbell's (1979) hypothetical outcomes appear in Figure 13-14. Can you interpret each set of findings pictured there?

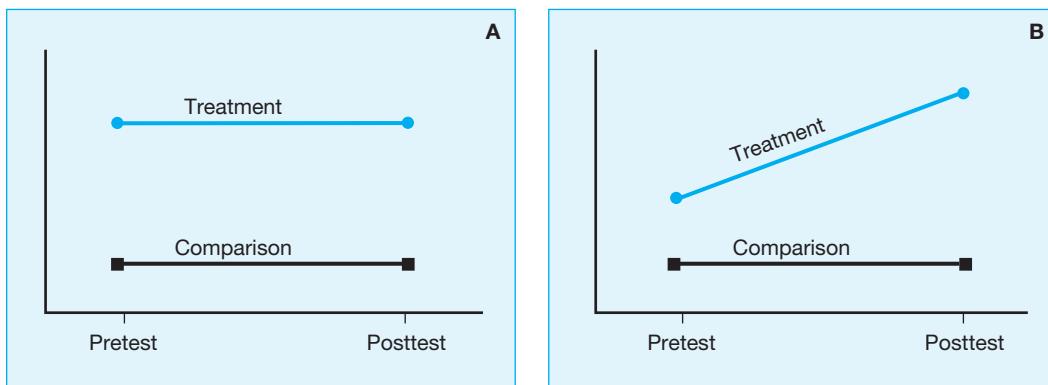


FIGURE 13-13 Two Possible Outcomes in a Nonequivalent Group Design.

Source: *Quasi-Experimentation: Design and Analysis*, by Thomas D. Cook and Donald T. Campbell, 1979, New York: Houghton-Mifflin. Reprinted by permission.

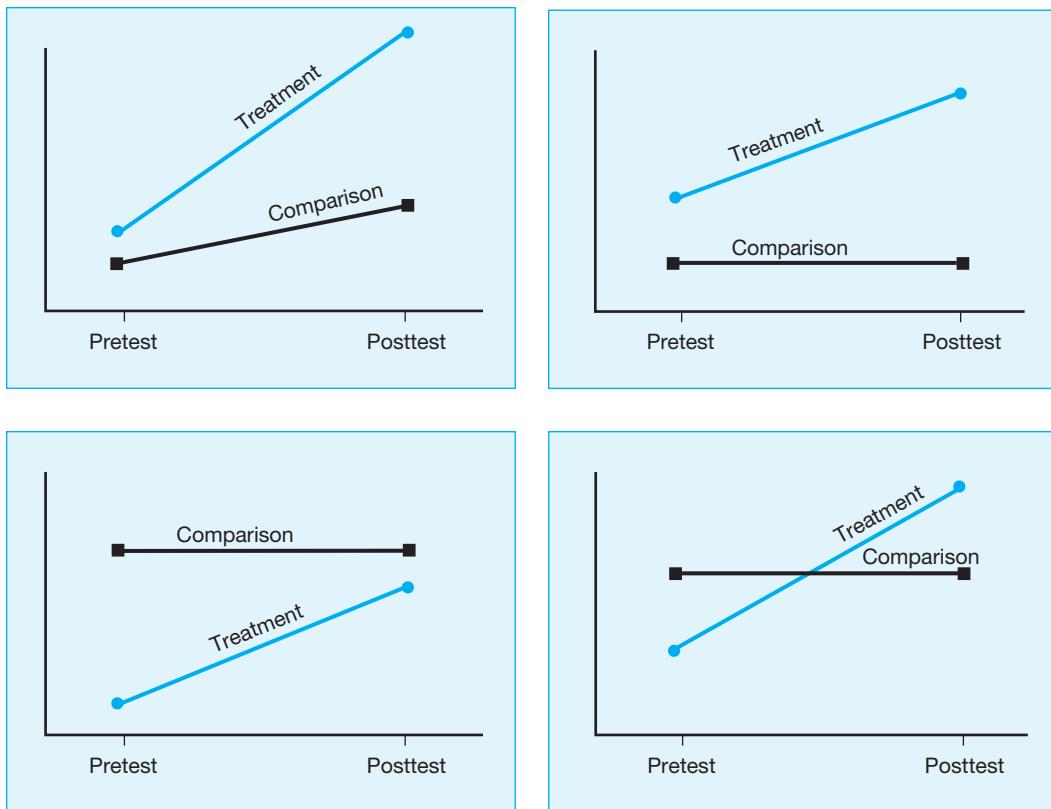


FIGURE 13-14 Several Additional Outcomes in a Nonequivalent Group Design.

Source: *Quasi-Experimentation: Design and Analysis*, by Thomas D. Cook and Donald T. Campbell, 1979, New York: Houghton-Mifflin and Company. Reprinted by permission.

Thus far, our discussion of this design has seemed similar to that of true experimental designs. What is different about quasi-experimental designs? The most important point to remember is that quasi-experimental designs are more plagued by threats to internal validity. Because you have not used random assignment, your interpretation of the findings must be cautious. Cook and Campbell (1979) isolated four threats to internal validity that are not controlled in the nonequivalent group design. We will list these threats only briefly because they appeared in Chapter 8. First, *maturity* is a potential problem. Because the groups begin as unequal, there is a greater potential that results such as those shown in Figure 13-13B might be due to differential maturation of the groups rather than to the IV. Second, we must consider *instrumentation* in the nonequivalent group design. For example, if we demonstrate nonequivalence of our participants by using a scale during the pretest, we must worry about whether the scale is uniform—are the units of measurement equal throughout the scale? *Statistical regression* is the third internal validity threat present in the nonequivalent group design. Regression is particularly likely to be a problem if we select extreme scorers on the basis of our pretest. Finally, we must consider the threat to internal validity of an *interaction between selection and history*. If some local event differentially affected our treatment and comparison groups, we would have a problem.

In conclusion, the nonequivalent group design is a strong quasi-experimental design. Its strength lies in the fact that “it provides an approximation to the experimental design and that, with care, it can support causal inference” (Hedrick et al., 1993, p. 62). Of course, we must be aware that the threat of confounds is higher than it is in the true experimental designs. Hedrick et al. (1993) warned that “throughout both the planning and execution phases of an applied research project, researchers must keep their eyes open to identify potential rival explanations for their results” (p. 64). Often researchers who use quasi-experimental designs must address potential alternative hypotheses in their research reports.

Interrupted Time-Series Design Another quasi-experimental design, the **interrupted time-series design**, involves measuring a group of participants repeatedly over time (the time series), introducing a treatment (the interruption), and measuring the participants repeatedly again (more of the time series). Look at Figure 13-15 to see a graphic portrayal of an interrupted time-series design. We should make an important point about Figure 13-15: There is nothing magical about using five observations before (O_1 – O_5) and after (O_6 – O_{10}) the treatment. Any number of observations large enough to establish a pattern can be used (Campbell & Stanley, 1966, showed four before and after; Cook & Campbell, 1979, showed five; Hedrick et al., 1993, showed six before and five after). As you can probably guess, the idea behind an interrupted time-series design is to look for changes in the trend of the data before and after the treatment is applied. Thus, the interrupted time-series design is similar to an A-B design. A change in trend could be shown by a change in the *level* of the behavior (see Figure 13-16A), a change in the *rate* (slope) of the pattern of behavior (see Figure 13-16B), or both (see Figure 13-16C).

Interrupted time-series designs have been used for quite some time. Campbell and Stanley (1966) referred to their use in much of the classical research of nineteenth-century biology and physical science. Cook and Campbell (1979) cited a representative 1924 study dealing with the effects of moving from a 10-hour to an 8-hour workday in London. Hedrick and Shipman (1988) used an interrupted time-series design to assess the impact of the 1981 Omnibus Budget Reconciliation Act (OBRA), which tightened eligibility requirements for Aid to Families with Dependent Children (AFDC) assistance. As shown in Figure 13-17, the immediate impact of this legislation was to lessen the number of cases handled by about 200,000; however, the number of cases after the change continued to climb at about the same slope it had before the change. Thus, the tightened eligibility requirements seemed to lower the *level* of the caseload but not its *rate*.

Interrupted time-series design A quasi-experimental design, involving a single group of participants, that includes repeated pretreatment measures, an applied treatment, and repeated post-treatment measures.



Review the threats to internal validity summarized in Chapter 8. Which threat would seem to create the greatest potential problem for the interrupted time-series design?

O_1	O_2	O_3	O_4	O_5	X	O_6	O_7	O_8	O_9	O_{10}
-------	-------	-------	-------	-------	---	-------	-------	-------	-------	----------

FIGURE 13-15 An Interrupted Time-Series Design.

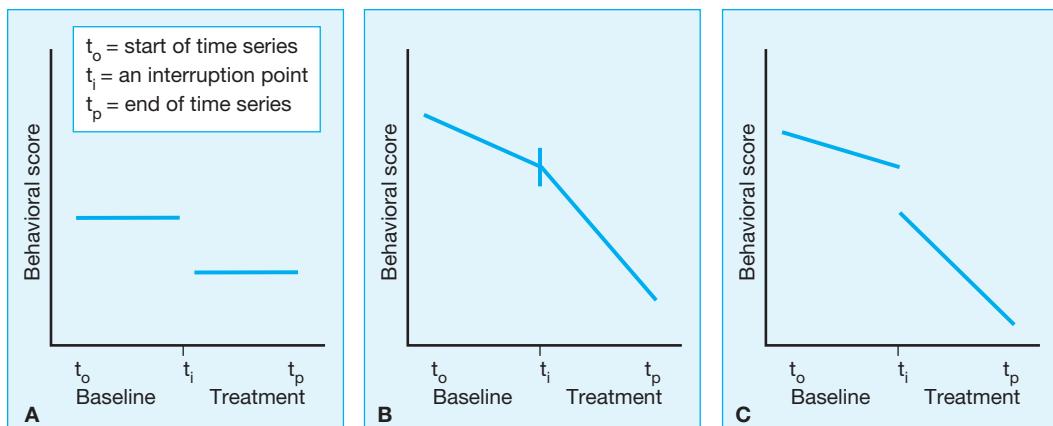


FIGURE 13-16 Potential Changes in Trend in a Time-Series Design. **A.** Change in level, no change in rate. **B.** No change in level, change in rate. **C.** Change in level, change in rate.

Source: Portions of Figure 1 from "Time-Series Analysis in Operant Research," by R. R. Jones, R. S. Vaught, and M. Weinrott, 1977, *Journal of Applied Behavior Analysis*, 10, pp. 151–166.

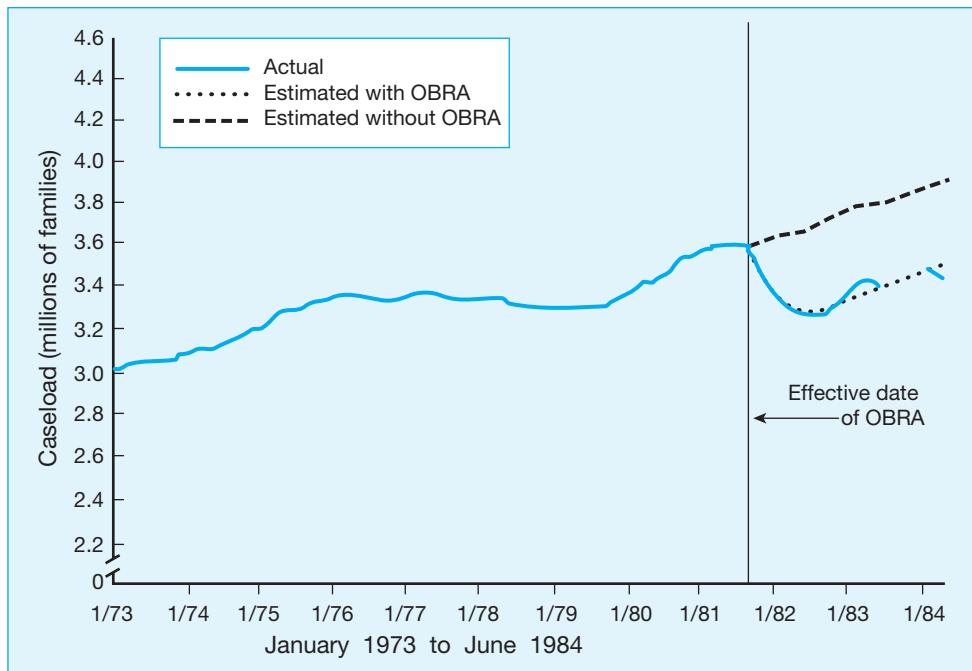


FIGURE 13-17 Impact of Tightened AFDC Requirements on Caseload.

Source: "Multiple Questions Require Multiple Designs: An Evaluation of the 1981 Changes to the AFDC Program," by T. E. Hedrick and S. L. Shipman, *Evaluation Review*, 12, p. 438. Copyright © 1988 Sage Publications, Inc. Reprinted by permission of Sage Publications, Inc.

According to Cook and Campbell (1979), the main threat to most interrupted time-series designs is *history*. One of the primary features of the interrupted time-series design is the passage of time needed to take many different measurements. This time passage raises the possibility that changes in behavior *could* be due to some important event other than the treatment. Because of the time taken by repeated measurements, another potential threat to internal validity is *maturity*. Repeated pretesting does, however, allow for the assessment of any maturational trends: If scores change at the same rate before and after the treatment, the change is due to maturation. *Instrumentation* could be a problem if record-keeping or scoring procedures change over the course of time. Such a change, of course, would violate the principles of control in any experiment, not just an interrupted time-series design.

Although the interrupted time-series design can control for some of the internal validity threats, we still face the potential problem of history. This threat to internal validity is usually handled in one of three manners. First, Cook and Campbell (1979) advised frequent testing intervals. For example, if you test participants on a weekly rather than monthly, quarterly, or yearly basis, the probability of a major event occurring during the time period between the last pretest and the treatment is low. In addition, if you keep careful records of any possible effect-causing events that occur during the quasi-experiment, it would be a simple matter to discern whether any occurred at the critical period when you administered the treatment. This first approach to controlling history is probably the most widely used because of its ease and the drawbacks involved with the next two solutions.

The next solution to the history threat is to include a comparison (control) group that does not receive the treatment. Such a design appears in Figure 13-18. As you can see, the comparison group receives the same number of measurements at the same times as the treatment (experimental) group. Thus, if any important historical event occurs at the time the experimental group receives the treatment, the comparison group would have the same experience and show the same effect. The only problem with this solution is that the comparison group would most likely be a nonequivalent group because the groups were not randomly assigned. This nonequivalence would put us back in the situation of attempting to control for that difference, with the associated problems we covered in the previous section of this chapter.

The third possible solution to the history problem is probably the best solution, but it is not always possible to do. In essence, this solution involves using an A-B-A format within the interrupted time-series design. The problems, of course, are those that we mentioned earlier in the chapter when dealing with the A-B-A design. Most important, it may not be possible to "undo" the treatment. Once a treatment has been applied, it is not always reversible. Also, if we halt an experiment in the A stage, we are leaving our participants in a nontreatment stage, which may have negative consequences. Hedrick et al. (1993) presented the results of an unintentional interrupted time-series design in an A-B-A format. In 1966 the federal government passed the Highway Safety Act, including a provision that mandated helmets for motorcyclists.

O ₁	O ₂	O ₃	O ₄	O ₅	X	O ₆	O ₇	O ₈	O ₉	O ₁₀
O ₁	O ₂	O ₃	O ₄	O ₅		O ₆	O ₇	O ₈	O ₉	O ₁₀

FIGURE 13-18 An Interrupted Time-Series Design with Control Group.

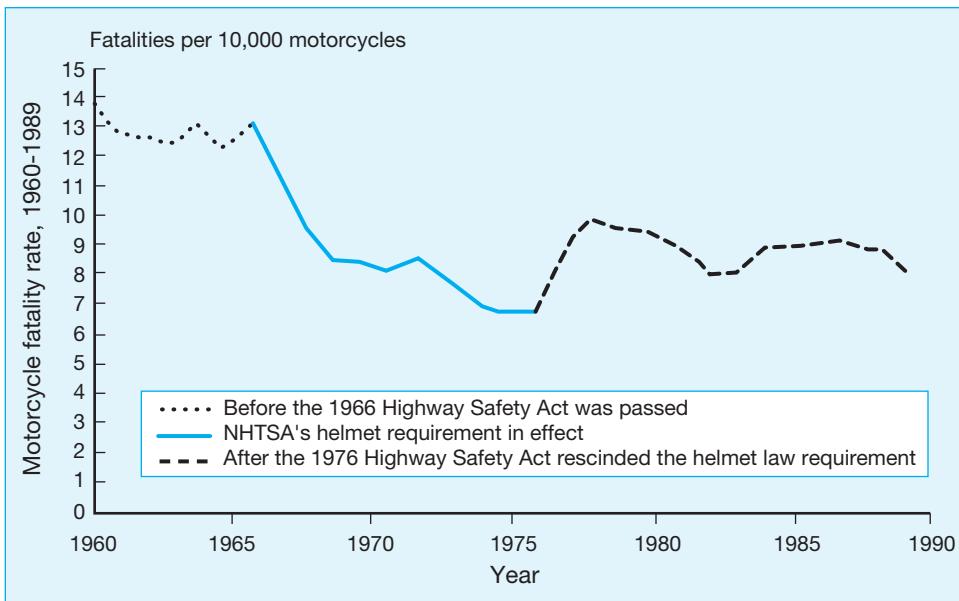


FIGURE 13-19 Effects of Mandatory Helmet Laws and Their Subsequent Repeal on Motorcycle Fatalities.

Source: Motorcycle Helmet Laws Save Lives and Reduce Costs to Society (GAO/RCED-91-170, July), Washington, DC. From *Applied Research Design: A Practical Guide*, by T. E. Hedrick, L. Bickman, and D. J. Rog, 1993, Newbury Park, CA: Sage, p. ??.

In the late 1970s states began to repeal the helmet laws as a result of the pressure being applied by individuals concerned with freedom of choice. If we examine motorcycle fatality rates over many years, we have an A (no restrictions), B (helmet laws), A (fewer restrictions) format for an interrupted time-series design. Figure 13-19 shows a graph presented by Hedrick et al. (1993). Because of the drop in fatalities after the law was passed and the rise in fatalities after the law was repealed in some states, it seems straightforward to derive a cause-and-effect relation from these data. Although this type of design allows for a convincing conclusion, again we must point out that the circumstances that created it are unusual and would be difficult, if not impossible, to recreate in many typical quasi-experimental situations.

In summary, the interrupted time-series design has the ability to uncover cause-and-effect relations. You must be especially careful of history effects when using this design; however, frequent testing can reduce this threat. The interrupted time-series design is particularly helpful when you are dealing with applied types of problems such as therapeutic treatment or in educational settings.

■ REVIEW SUMMARY

1. **Quasi-experimental designs** are identical to true experimental designs except that participants are not randomly assigned to groups. Thus, our research groups may not be equal before the experiment, which can cause problems in drawing clear conclusions.
2. Unlike the case in ex post facto designs, we are able to control the IV in a quasi-experimental design.

3. There are many situations in which the impossibility of random assignment makes quasi-experimentation necessary.
4. The **nonequivalent group design** involves comparing two groups—one of which receives the IV and a comparison group that does not receive the IV. The groups are non-equivalent because of the lack of random assignment.
5. In the nonequivalent group design, it is imperative to select a comparison group that is as similar as possible to the treatment group.
6. **Maturation, instrumentation, statistical regression**, and **selection-history interactions** are all threats to internal validity in the nonequivalent group design.
7. An **interrupted time-series design** involves measuring participants several times, introducing an IV, and then measuring the participants several more times.
8. **History** is the main threat to internal validity in the interrupted time-series design. It can be controlled by testing frequently, including a comparison group, or removing the treatment after it has been applied (if possible).

Check Your Progress

1. Differentiate between experimental designs, quasi-experimental designs, and ex post facto designs.
2. Give two reasons why you might choose to use a quasi-experimental design rather than an experimental design.
3. Match the design with the appropriate characteristics:

1. nonequivalent group design	A. typically has one group of participants
2. interrupted time-series design	B. has two groups of participants
	C. involves pretesting participants
	D. does not involve pretesting participants
	E. is prone to the internal validity threat of history
	F. is prone to several internal validity threats
4. What was the key to Geronimus's (1991) research that allowed her to conclude that the effects of teenage pregnancy are not as negative as typically thought?
5. We summarized two interrupted time-series analyses in the text: one dealing with changing AFDC requirements (Figure 13-17) and one dealing with changing motorcycle helmet laws (Figure 13-19). Why are we more certain about our conclusion in the case of the helmet laws than with the AFDC requirements?
6. If Prohibition (the outlawing of alcoholic beverages in the 1920s) were to be treated as an experiment to determine its effects on alcohol consumption, what design would this represent?
 - a. nonequivalent group design
 - b. single-case design
 - c. interrupted time-series design with control group
 - d. interrupted time-series design

■ Key Terms

Internal validity, 300	Interactions with selection, 305	A, 314
Confounding variables, 300	Experimental mortality, 305	B, 314
Extraneous variables, 300	Instrumentation, 305	A-B design, 314
Cause-and-effect relation, 300	Diffusion or imitation of treatments, 305	A-B-A design, 315
Random assignment, 303	Single-case experimental design, 308	A-B-A-B design, 316
Random selection, 304	Case-study approach, 309	Quasi-experimental design, 321
Selection, 304	Experimental analysis of behavior, 310	Nonequivalent group design, 323
History, 304	Baseline, 312	Interrupted time-series design, 327
Maturation, 304		
Testing, 304		
Statistical regression, 304		

■ Looking Ahead

At this point you may have finished planning, conducting, and analyzing your research project. Still, one task lies ahead—writing the research report, which is the culmination of your research effort. In the next chapter we will cover how researchers write their reports in the American Psychological Association's style.

Writing and Assembling an APA-Format Research Report

CHAPTER
14

What Is APA Format?

Sections of the APA-Format Paper

- Title Page • Abstract • Introduction • Method
- Results • Discussion • References • Appendix
- Author Note • Headings

Writing in APA Style

- General Guidelines • Grammatical Guidelines

APA Editorial Style

Preparing Your Manuscript

Student Views of Professional Activities

The final step in the scientific process is communicating your results to other people who are interested in your research topic. As you have seen in numerous police shows, detectives also have to spend time writing their reports. Their purpose is no different from ours—communicating what has happened to a particular audience. Sherlock Holmes knew the importance of communicating clearly when he said to Watson, “Never trust to general impressions, my boy, but concentrate yourself upon details” (Doyle, 1927, p. 197). We have hinted at writing and organizing various parts of a manuscript in American Psychological Association (APA) format throughout several chapters of this book. In this chapter we will cover all the aspects of writing a research report in APA format. We will use a student research project as our example, tracing it from the research stage through manuscript preparation, to publication. In this chapter we will introduce Allison Dickson, Traci Giuliano, James Morris, and Keri Cass’s (2001) research project concerning the effects of race of performer and type of music on ratings of a musical performer. Dickson, Morris, and Cass were students at Southwestern University in Georgetown, Texas, and Giuliano was their supervising professor. We include this research as an example not to intimidate you, but to show you what is possible at the undergraduate level.

What Is APA Format?

Although several styles exist for the preparation of reports in various academic disciplines, psychologists developed their own format to meet their specific needs. Because the American Psychological Association originally developed this style, it is often referred to as **APA format**. In Chapter 1 we saw that in the 1920s University of Wisconsin psychologist Joseph Jastrow found a need to bring uniformity to the publication of research articles in our field. The lack of a set model or pattern had resulted in published research reports that were nearly impossible to compare. In addition to the fact that there was no prescribed order for the presentation of information, there was no consistent manner for describing one’s methodology, procedure, or data analysis. In short, the order

APA format Accepted American Psychological Association form for preparing reports describing psychological research.

and manner of presentation varied from one article to the next. The APA format for preparing papers was developed to bring order to this state of chaos.

The particular form for preparing APA format papers discussed here is found in the fifth edition of the *Publication Manual of the American Psychological Association* (American Psychological Association, 2001). The APA *Publication Manual (PM)* has changed and evolved since it was first published in 1952. Many of the changes that have occurred over the years, as well as much of the actual format itself, were implemented to assist printers—after all, APA format was adopted to help make the publication of journal articles more uniform. Although many earlier APA-format matters had the printer in mind, the latest edition of the APA *PM* clearly shows the influence and importance of the computer. For example, the earlier requirement of large 1½-inch margins has been replaced by more computer-friendly 1-inch margins. Likewise, the reference section has reverted back to hanging indentations (the first line begins at the margin and subsequent lines are indented) from the third edition because computers and word processing programs can easily deal with hanging indents. In addition to being computer friendly, the general

layout and design of the APA-format paper are reader friendly. For example, using separate, specifically designated sections allows the reader to know exactly which part of the project is being discussed. Authors use **headings** to divide the APA-format paper into sections to help the reader understand the paper's organization.

Before we get started, we have to make an important point: This chapter is not meant to substitute for the *PM* (APA, 2001). There is simply no way for us to condense a 439-page book into a single chapter for this text. You should buy a *PM* and think of it as an investment in your future. As well as using it in this class, you will use it when you write term papers in other psychology classes and for any other writing you will do if you go to graduate school in psychology. In addition, other academic disciplines are beginning to adopt APA style for their writing assignments. One of your authors has had colleagues in education, political science, and communication disorders borrow his *PM* for pointers on formatting and style.

Headings Titles for various sections of a psychology paper that are designed to help the reader understand the outline and importance of the parts of the paper.

Sections of the APA-Format Paper

The major components of an APA-format paper are, in order:

1. Title page
2. Abstract
3. Introduction
4. Method section
5. Results section
6. Discussion section
7. References
8. Appendixes (if any)
9. Author note
10. Tables (if any)
11. Figures (if any)

As we look at the various sections of the APA-format paper, we will refer to the Dickson, Giuliano, Morris, and Cass manuscript as an example (Figures 14-1 through 14-16; *M* indicates manuscript, and the number after *M* represents page order). In addition, you can refer to the published article (Dickson, Giuliano, Morris, & Cass, 2001) displayed in Figures 14-17 through 14-21 (*JA* indicates journal article) to see how a manuscript is reformatted as a journal article. (After our initial citation, the correct format for citing this manuscript is Dickson et al., 2001; we will refer to it using only Dickson's name in order to conserve space.)

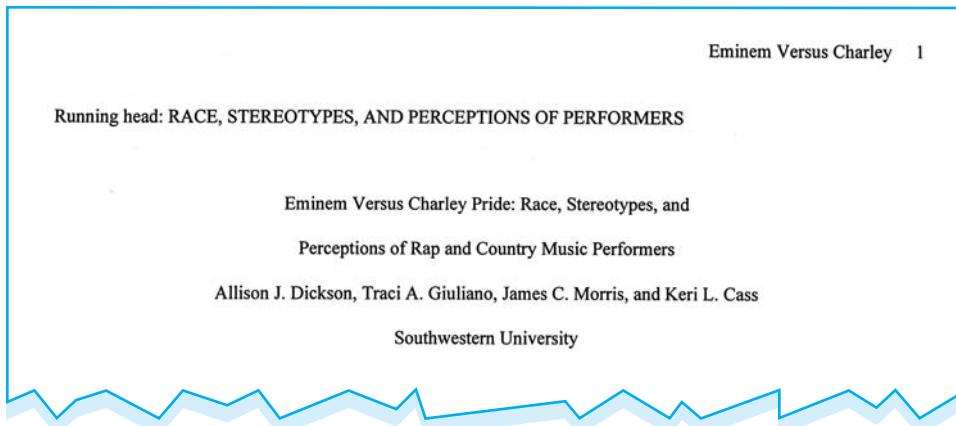


FIGURE 14-1 (M1). Title page.

Title Page

The title page of Dickson's manuscript appears in Figure 14-1. The PM contains information about the **title page** on pages 10–12 and 296–298. The important features are the manuscript page header, page number, running head, author(s), and affiliation(s). The **manuscript page header**, which consists of the first two or three words of the title, is used to identify the pages of a particular manuscript should they become separated. It appears in the upper right-hand portion of each page two lines above the 1-inch margin. The page header appears either five spaces to the left of the page number or two lines above the page number. Neither the page number nor the page header should extend beyond the margin on the right side of the page. The manuscript page header and page number appear on *all* pages of the paper, *except* the figures.

Two lines below the manuscript page header, flush against the left margin, we find the **running head**. The running head, which is a condensed title, will be printed at the top of alternate pages when your paper is published in a journal. When you type the running head on the title page, it appears in all-capital letters. Also, note that the *h* in running head is not capitalized. The running head “should be a maximum of 50 characters, including letters, punctuation, and spaces between words” (APA, 2001, p. 12). As you can see in Figure 14-1, Dickson used different phrases for the manuscript page header and the running head. Although the same phrase might be a logical choice for both elements, APA style does not specify that they must be identical. The manuscript page header may not make sense, but the running head should communicate information about the contents of your manuscript because it would be printed in a published article (see Figure 14-18, p. 366).

The title of the paper, which is centered, may begin six or eight lines below the page number. Capitalize the first word and all major words (4 or more letters, according to the PM) of the

Title page The first page of an APA-format paper. It includes the manuscript page header, the running head, the manuscript's title, and the name(s) of the author(s) and their affiliation(s).

Manuscript page header The first two or three words of the report's title. Appears with the page number on each page of the research report.

Running head A condensed title that is printed at the top of alternate pages of a published article.

title. Your title should summarize clearly and simply the nature of the research you are reporting, but it should not be overly long. APA's recommended length for the title is from 10 to 12 words. If your title is too long to fit on one line, you can type it on two lines. You should be especially careful when you choose a title for your paper. Although it is often tempting to compose a title that is cute or catchy, such titles are often failures at communicating anything about the article's content. Keep in mind that many people will read your title and will use it to decide whether to read the entire paper. Dickson's title is slightly longer than the APA's guideline; the subtitle fully communicates that the paper deals with the variables of race, stereotypes, and perceptions of musical performers. The title is slightly longer than the recommended length because of the four-word "main" title—although it does not communicate much about the content of the article, it is certainly attention getting. Your instructor (or journal editor) may give you some leeway with the 10- to 12-word recommendation, as in this case.

The name of the author is double-spaced below the title. The author's institutional affiliation is double-spaced below the author's name. If there is more than one author, the authors' names appear in order of importance of their contributions to the research project. If authors are from the same institution, all the names are on one line, if they will fit. Dickson's final manuscript included an additional author. The class project was a group effort conducted by Dickson, Morris, and Cass for their research methods course, so they were the original authors. Giuliano was the professor of the class and helped them develop their ideas, supervised the research project, assisted in the statistical analysis and interpretation, and participated with Dickson in cowriting the manuscript for submission after the class was finished. Morris and Cass did not participate in rewriting the manuscript after the class. Thus, although the research originated as a student group project, by the time it was published, it had become a team effort among three students and a faculty member. Sometimes it is difficult to decide whether an individual merits being listed as an author on a publication. According to the *PM*, "Authorship encompasses not only those who do the actual writing but also those who have made substantial scientific contributions to a study" (p. 6). This guideline is somewhat vague and results in people having to make judgment calls about authorship.

Abstract

Level 1 heading A centered section title in which the first letters of major words are capitalized. Occupies a line by itself.

Figure 14-2 shows the abstract page. Note, once again, the placement of the manuscript page header and page number. The word "Abstract" is centered and appears two lines below the page number. A centered section title in which the first letters of major words are capitalized is designated as a **Level 1 heading**. Remember that headings demarcate sections of the APA-format paper. (Full information explaining different levels of headings appears on pp. 289–290 in the *PM* and on pp. 363–364 in this book.)



Imagine that you have just picked up a journal issue and are skimming through it to see whether you might be interested in reading any of the articles. What feature of an article will you use as a guide in deciding whether to read it? What information would you use in making this decision?

Abstract

The present study explored the effects of stereotype deviation in the music industry on people's perceptions of performers. One hundred college students (48 men, 52 women) examined a profile of a fictitious performer containing a picture, a brief biography, and a lyric sample. As part of a 2-way between-subjects design, participants made judgments about either a Black or a White musician who performed either rap or country music. The results showed that a Black rap performer was rated more favorably than a Black country performer, and a White country performer was rated more favorably than a White rap performer. Consistent with predictions, people who violate societal expectations are judged more harshly than are people who conform to societal expectations, particularly in cases involving strong preexisting racial stereotypes.

FIGURE 14-2 (M2). Abstract.

Of course, the title of this section answers the first question; you will use the abstract as a guide to determine whether you want to read the complete article. The **abstract** of an experimental report consists of a brief (up to 120 words), one-paragraph description of the research presented in your paper. In order to help you make an educated decision about pursuing an article in depth, the paragraph comprising the abstract should include a description of the intent and conduct (including participants and method) of your project, the results you obtained, and the project's implications or applications. Note that the abstract is typed in block form; there is no indentation on the first line.

The *PM* states that an abstract for an experimental study should contain information about the problem, participants, experimental method, findings, and conclusions.

Abstract A brief description of the research that is presented in an APA-format paper.



Can you find information about each of the five topics (problem, participants, method, findings, conclusion) in the abstract shown in Figure 14-2?

Here's where we found the relevant information:

Problem: lines 1–2

Participants: line 2

Experimental method: lines 2–5

Findings: lines 5–6

Conclusions: lines 6–8

Did you find all the necessary information? Writing a good abstract is challenging because of the large amount of information that must be covered in little space. You can find more information about abstracts on pages 12–15 and 298 of the *PM*.

Introduction

Introduction The first major section of the APA-format paper. Contains the thesis statement, review of relevant literature, and experimental hypothesis.

Your **introduction** section begins on page 3 of your report. Notice that the word *Introduction* does not appear as the Level 1 heading that begins this section. Rather, you repeat the title from page 1. Be sure that the title is *exactly* the same in both places. Figure 14-3 shows the first page of Dickson's introduction section (Figure 14-4 on page 340 shows the second page of the introduction); the concluding page of the introduction section appears in Figure 14-5 on page 341.

We like to use the analogy of a funnel to describe a good introduction section. Like the introduction, a funnel starts off broad, then narrows to a specific focus. At the end the focus is very specific, and in the case of the introduction section it leads to a logical experimental question: your experiment. In Dickson's example, the first paragraph is broad, establishing the race and genre of musicians as the variables of interest. The second and third paragraphs provide basic background information about racial stereotypes. The fourth and fifth paragraphs begin to narrow the funnel. Here, you read about some specific effects of stereotypes and race associated with music, as well as the effects of stereotype inconsistencies. Finally, the last paragraph lays out the specific experimental hypothesis of this research.



Can you identify the thesis statement in Dickson's introduction?

Thesis statement A statement of the general research topic of interest and the perceived relation of the relevant variables in that area.

The first sentence on page 5 (Figure 14-5) forms the thesis statement for the student manuscript. The **thesis statement** should indicate the general topic in which you are interested and your general view of the relation of the relevant variables in that area. Beginning with "we expected that," you see the specific predictions made for the performance of participants rating musical artists under stereotype-consistent conditions compared to stereotype-inconsistent conditions. Some writers prefer to present the thesis statement in the opening paragraph of the introduction. The exact location is not important as long as you include your thesis statement somewhere in the introduction. The *PM* provides more information concerning introductions on pp. 15–17.

Other than presenting the thesis statement, the introduction reports the results of previous research that support or refute your thesis statement. This material represents the outcome of your literature review. Give careful thought to its presentation; good organization and a logical flow of ideas are important ingredients. The introduction is similar to telling a story to someone who may not be familiar with what you are talking about; you must be clear and not make unwarranted assumptions. First, you establish the big picture. Then, you begin

Eminem Versus Charley 3

Eminem Versus Charley Pride: Race, Stereotypes, and
Perceptions of Rap and Country Music Performers

What do Eminem and Charley Pride have in common? Perhaps the connection these two performers share is subtle, but they are in fact quite similar in at least one aspect of their careers. Both Charley Pride, a Black country music performer, and Eminem, a White rap performer, deviate from social expectations that are a part of the music industry. Specifically, these two musicians defy cultural stereotypes by performing types of music that are not typically associated with their race.

Racial stereotypes exist in most individuals, and they can influence subsequent judgments made by a perceiver (Devine, 1989; Dovidio, Evans, & Tyler, 1986; Gaertner & McLaughlin, 1983). For example, Gaertner and McLaughlin (1983) studied the effect of racial stereotypes on perceptions and found that White students responded faster to positive stereotyped words (e.g., *smart*) when the words followed the race *White* rather than *Black*. In addition, Sagar and Schofield (1980) examined the perceptions made by sixth-grade boys about ambiguous behavior. They found that both Black and White boys construed ambiguously aggressive behaviors (such as one child bumping into another) as being more threatening if the actions were performed by a Black boy rather than a White boy. Most people today would not be likely to openly express racist beliefs, but the results of the above studies support the aversive racism perspective, which suggests that subtle and indirect forms of racism persist in society today (Dovidio & Gaertner, 1991; Gaertner & Dovidio, 1986). That is, although current cultural values emphasize fairness and racial equality, White individuals have a historic tradition of having negative beliefs concerning Blacks and other minority groups (Dovidio, Brigham, Johnson, & Gaertner, 1996). Consequently, racial stereotypes continue to exist and to influence interactions among individuals in society, but perhaps in more subtle ways.

Because stereotypes can influence the judgments and behaviors of perceivers, deviations from a stereotype should have similar effects. In general, the expectations of the perceiver and the extent to which these expectations are confirmed or disconfirmed can influence judgments. When behavior only slightly varies from expectations, the difference might not be noticed, but perceivers often magnify the discrepancy when actions differ significantly from expectations. This phenomenon is known as the contrast effect (Brehm, Kassin, & Fein, 1999). In fact, a person who displays behavior inconsistent with societal expectations is often evaluated more extremely than is a person who

FIGURE 14-3 (M3). First page of introduction.

behaves consistently with expectations (Knight, Giuliano, & Sanchez-Ross, 2001). Jackson, Sullivan, and Hodge (1993) examined the effects of describing stereotype-consistent or stereotype-inconsistent behavior of Black out-group targets and White in-group targets on social evaluations made by participants who assessed a college application. They found that people who deviate from a norm are judged more extremely than if they behave as the norm dictates. Specifically, stereotype-inconsistent Black applicants with strong credentials were evaluated more favorably than were strong White applicants, and stereotype-inconsistent White applicants with weak credentials were evaluated less favorably than were weak Black applicants.

Both stereotypes and stereotype-inconsistent behavior affect the evaluations individuals make about other people in a variety of social interactions. Fried (1996, 1999) studied biased reactions involving the music industry and found that individuals had very different reactions to music labeled as *rap* or *country* despite the fact that the song lyrics were exactly the same. In two studies, she found that people generally considered rap music to be more violent and more offensive than country music. Furthermore, a folk song that was presented as being performed by a Black artist was judged more negatively than the very same song when it was presented as being performed by a White artist (Fried, 1996). Fried (1999) attributed her results to stereotypes in that rap is usually associated with Black culture whereas country music is often thought of as being a part of White culture. By priming a Black stereotype with the use of the label *rap*, it is possible that individuals apply negative stereotypes that have been shown to be associated with African Americans (Brigham, 1971). Therefore, racial stereotypes can impact evaluations of music performers (Fried, 1996, 1999).

According to Fiske and Taylor (1991), long-held stereotypes are not easily altered, but modification of stereotypes may begin with a divergence by stereotyped individuals or groups. Previous research has examined the effects of stereotypes and stereotype deviation on people's evaluations of other individuals (Jackson et al., 1993; Knight et al., 2001). The present study attempted to integrate and expand on these concepts in relation to the music industry. Specifically, whereas Fried (1996, 1999) examined the effect of either the race of the performer or the labeled genre of music of a song on evaluations about the music itself, the design of the present experiment explored the interactive effects of the race of the performer and the genre of music on participants' evaluations of the performer. In doing so, we explored the difference between perceptions of persons who adhere to social expectations versus persons who deviate from the stereotype.

FIGURE 14-4 (M4). Second page of introduction.

Consistent with previous research (Jackson et al., 1993; Knight et al., 2001), we expected that participants would judge performers who behave consistently with social norms (i.e., Black rap artists and White country performers) more favorably than performers who deviate from societal expectations (i.e., White rap artists and Black country performers). That is, because country music is associated with White culture, a Black country performer does not exhibit behavior consistent with this stereotype and, as a result, this performer should elicit negative judgments. The same reaction should occur with a White rap artist because his or her behavior is inconsistent with the stereotype that rap is predominantly a part of Black culture.

Method

Participants

Data were collected from 100 undergraduates (48 men, 52 women) at Southwestern University, a small liberal arts college in the Southwest. Demographically, the university is composed primarily of White, middle- to upper middle-class students; as such, the current sample (which was representative of the campus at large) consisted almost exclusively of White students. Participant volunteers ranged in age from 18 to 26 years ($M = 19.77$ years). Data from four participants were excluded from the analysis because these participants either failed to follow instructions or they did not pass the manipulation check. Specifically, these participants were unable to identify the race of the performer and genre of his music that was presented in their survey packet.

Design and Procedure

The present study used a 2 (race of performer: Black or White) \times 2 (genre of music: rap or country) between-subjects design to explore the effect of the race of a performer and the genre of music on perceptions of the performer. We recruited participants from various locations on campus and asked them to contribute to an investigation exploring “people’s perceptions of music.” Once they agreed, participants viewed a picture of a male performer, read a brief biography about him, and read a lyric sample of his music. Next, they completed a survey in which they made judgments about the performer and his music, and they responded to filler questions concerning their taste in music in general to corroborate the cover story. Each participant was randomly assigned to one of four experimental conditions and read a profile of either a Black rap artist, a White rap artist, a Black country performer, or a White country performer. Measures of the primary dependent variable (i.e., how favorably participants rated the performers) were embedded among filler questions in the survey. Following completion of the survey and a brief

FIGURE 14-5 (M5). Last page of introduction; first two method subsections.

to fill in the details. Your ultimate goal is to lead the reader to a description of what you plan to do in your experiment. This description usually appears at the end of the introduction section. It is here that you will state your experimental hypothesis.

Citation A notation in text that a particular reference was used. The citation provides the name(s) of the author(s) and date of the work's publication.

Reference A full bibliographic record of any work cited in the text of a psychological paper.

Reference section A complete listing of all the references cited in a psychological paper.

Although we cannot teach you *how* to write in the limited space available, we can give you a few pointers to assist you in preparing your introduction section. Note that every fact-based statement is supported by a **citation** to a **reference**. If you wish to locate one of the references cited, you can find the necessary information in the **reference section** (also known as the *reference list*) at the end of the paper (see Figures 14-14 and 14-15, pp. 358–359). As you can see from Dickson's introduction, when citing references in the text of an APA-format paper, you use the last name of the author(s) and the year of publication. Such citations can take one of two forms. In the first form, the authors are either the subject or the object in a sentence in the text. In such cases, only the date of the publication appears in parentheses. This type of citation takes the following form:

Jackson, Sullivan, and Hodge (1993) examined the effects . . .

In the second form of citation, you may wish to cite a reference only as support for a statement you have made. In this instance you would include the author(s) and the year of publication *inside* the parentheses. This type of citation takes the following form:

This phenomenon is known as the contrast effect (Brehm, Kassin, & Fein, 1999).

There are two additional considerations for such citations. When there are multiple authors, the name of the last author is preceded by the ampersand (&) sign, *not* the word *and*. When you cite more than one study within parentheses, you should arrange the various citations alphabetically by the first author's last name.



How do you cite two papers published in the same year by exactly the same authors? How can you distinguish between these papers when you cite them in your research report?

To report papers published in the same year, simply place a lowercase *a* or *b* after the date of each reference in question. Thus, if you had two 2000 references by Smith and Jones, you would cite them as follows:

Smith and Jones (2000a, 2000b)

or

(Smith & Jones, 2000a, 2000b)

The *a* and *b* designations will *also* appear as part of the complete reference of the article in the reference section (see the section on references later in this chapter).

When a citation includes three to five authors, include each author's last name in the first citation; subsequent citations include only the first author's last name followed by et al. (Latin

for “and others”) and the date. As shown in paragraphs 3 and 5 of Dickson’s introduction, this rule creates the following citations:

- Jackson, Sullivan, and Hodge (1993) [first citation, paragraph 3]
 (Jackson et al., 1993) [second citation, paragraph 5]

If a citation includes six or more authors, *all* citations (including the first) consist of the first author’s last name followed by et al. and the date. In the reference section, you include the first six names followed by et al. The *PM* provides more information about reference citations in text on pages 207–214.

A second pointer on writing in APA style is that the use of unbiased language is imperative. **Unbiased language** is language that does not state or imply a prejudice toward any individual or group. According to the *PM*:

APA is committed both to science and to the fair treatment of individuals and groups, and this policy requires authors of APA publications to avoid perpetuating demeaning attitudes and biased assumptions about people in their writing. Constructions that might imply bias against persons on the basis of gender, sexual orientation, racial or ethnic group, disability, or age should be avoided. Scientific writing should be free of implied or irrelevant evaluation of the group or groups being studied.

Long-standing cultural practice can exert a powerful influence over even the most conscientious author. Just as you have learned to check what you write for spelling, grammar, and wordiness, practice reading over your work for bias. You can test your writing for implied evaluation by reading it while (a) substituting your own group for the group or groups you are discussing or (b) imagining you are a member of the group you are discussing. (APA, 2001, pp. 61–62)

Once you have presented your thesis statement, reviewed the relevant literature, and stated your experimental hypothesis, you are ready to tell the readers how you conducted your research. We turn to this topic next.

Method

The objective of the **method section** is to provide sufficient detail about your experiment to enable readers to evaluate its appropriateness or to replicate your study should they desire. The method section is typically made up of three subsections: *participants*, *apparatus* (also designated *materials* or *testing instruments*), and *procedure*. Note that “Method” is a Level 1 section head. You do not begin the method section on a new page if there is room on the previous page. There is no break between the introduction, method, results, and discussion sections of an APA-format paper. You begin on a new sheet of paper only if a heading falls on the last line of a page; if it does, move it to the next page and begin the section there.

Participants (Subjects) The **participants subsection** enumerates and describes the experimental participants. Figure 14-5 shows the first subsection of Dickson’s method section; *subjects* is a permissible label if an experimenter uses animals in the experiment. (Note that *Participants* is a **Level 3 heading**.)

Unbiased language

Language that does not display prejudice toward an individual or group.

Method section The second major section of the APA-format paper. Contains information about the participants; the apparatus, materials, and testing instrument(s); and the procedures used in the experiment.

Participants subsection The first subsection of the method section. Provides full information about the participants in the study.

Level 3 heading A section title that is left-margin justified, italicized, and has the first letter of each major word capitalized. Occupies a line by itself.

The subsection on participants answers three questions: Who participated in the study? How many participants were there? How were the participants selected? As you can see in Figure 14-5, Dickson gathered data from 100 students at Southwestern University. The students were almost evenly divided by sex; we see their age range, average age, and racial composition. She provided details about 4 participants excluded from the research for various reasons. Notice that she described the participants (and their important characteristics) in sufficient detail to allow a replication of the study. The only missing information is any inducement the participants received for being in the study; Dickson labeled them “volunteers,” but we do not know whether they received extra credit in a class, a monetary payment, or anything else. See pages 18–19 of the *PM* for more information about this subsection.

If you use animal subjects, your description must be detailed enough to allow another investigator to replicate your samples. In addition to providing sample selection information, you should indicate any special arrangements, such as housing conditions, dietary conditions, and so on. The following description of animal subjects, from a paper on evaluating rats’ behavior in a startle paradigm by Megan Kazlauskas, a student at Saint Anselm College in Manchester, New Hampshire, and Mark Kelland, her faculty advisor, is a good example of the type of information to include.

In this study we examined stargazer rats (homozygous stg/stg; stg group) and unaffected littermates (heterozygous stg/+; LM group) provided by Dr. Charles R. Ashby, Jr., of the Brookhaven National Laboratory (Upton, NY). Rat pups were bred in the vivarium and phenotyped at 14 days of age as stargazers or littermates, based upon the demonstration of stargazing behavior. After weaning, the rats were housed in pairs (stargazer with littermate), in rooms maintained at 25° C, 40% humidity, a 12-hr light/dark cycle (light, 0700–1900 hrs), with food and water available ad libitum. After being shipped to Saint Anselm College the rats were allowed to accommodate for 7 days, being maintained on a 12-hr light/dark cycle (light, 0800–2000 hrs), with food and water ad lib. All behavior measurements were performed during the light cycle, from 0800 to 1200 hrs. (Kazlauskas & Kelland, 1997, p. 92)

Materials subsection

The second subsection of the method section. When appropriate, contains information about materials other than the equipment used in the experiment.

Apparatus, Materials, or Testing Instruments Figure 14-6 shows the **materials subsection** of the method section. This subsection can have various names depending on what you use in your particular experiment; you should choose the term that best describes the elements you used in conducting your experiment. Dickson chose “Materials” because she created written information for her research. If you use items such as slides, pictures, videotapes, or paper-and-pencil tests that are not standardized, you would probably want to label this section *Materials*. For example, Dunham and Yandell (2005) developed a scale to measure artistic self-efficacy, so it was important to describe the scale items. They provided the following description:

Each question addresses one of the following drawing concepts or skills: contour lines, negative space, value, shading, texture, perspective, proportion, portraits, and composition. Example pictures were chosen to illustrate each skill. After the question topics were selected, they were then written on a basic reading level and did not contain any art jargon that non-artists would not understand. Two non-art professors from the Psychology Department reviewed the questions for clarity. The questions ranged in skill difficulty and were all paired with an example picture illustrating the skill in question. Each question is phrased similarly, beginning with “How well do you believe that you could . . .” and ending with “similar to the picture below.” All questions were answered with a 10 point Likert-type scale ranging from 1 (*cannot do at all*) and 10 (*certain can do*). The possible maximum score is 90, indicating high drawing self-efficacy, and the possible minimum score is 9, indicating low drawing self-efficacy. (Dunham & Yandell, 2005, p. 18)

Eminem Versus Charley 6

manipulation check that focused on the performer's race and the genre of music, we explained the purpose of the present study to the participants and asked them not to discuss it with anyone.

Materials

A three-page experimental packet, which ostensibly contained a survey about people's perceptions of music, was distributed along with an envelope. The first page contained a "performer profile" and included a color picture of a Black or White male performer, a brief biography about him indicating that he was either a rap or country performer, and a lyric sample from one of his songs. The subsequent pages contained the survey, which was used to measure participants' reactions to what they had seen and read on the previous page.

Each performer profile contained a biography that included the name of the performer (i.e., D.J. Jones), his hometown (Atlanta, Georgia), and a brief summary of his musical career (e.g., "D.J. has been singing since he was 14, and recently signed a record deal with a major country label. He will soon be on an international tour opening for a popular country artist"). To create the sample song lyrics, we slightly altered and combined two rock songs (May, 1973; May & Staffell, 1973). With the exception of the two manipulations, we used the same biography and song lyrics in each of the four experimental conditions. The first manipulation altered the genre of music (i.e., rap or country) that the artist performed. Specifically, the name of the performer and the type of music he performed was labeled beneath his picture, and this label coincided with the genre of music described in the biography. Next, a photograph manipulated the race of the performer so that each biography was accompanied by a picture of either a Black or a White man. To ensure that the only attribute differing between the Black and White performers was in fact race, we conducted a pilot test in a Research Methods class to match the Black and White performers on attractiveness and age. The participants in the pilot test were all White, which is representative of the ethnic composition of the participants in the present study. From a pool of 20 color photographs of nonfamous Black and White men selected from magazines, the two stimuli selected for the present study were most similar in perceived attractiveness and age.

Following the performer profile were the questions that assessed perceptions of the performer and his music, as well as the musical tastes of the participants. Embedded among demographic questions (e.g., sex and age) and other filler questions (e.g., "On average, how many CDs and/or tapes do you buy a month?") were the items that examined participants' perceptions of the performer. Specifically, participants rated on 7-point scales with endpoints

FIGURE 14-6 (M6). Materials subsection of method section.

Apparatus subsection

The second subsection of the method section. When appropriate, contains information about the equipment used in an experiment.

If you use equipment in the experiment, you should label this subsection the **apparatus subsection**. You should briefly describe the equipment used in the experiment and its function(s) in this subsection. If your equipment was commercially produced, mention the name of the manufacturer and the model number of the equipment. For example, in their study of memory for different types of images, Lucas et al. (2005) used computer equipment to present their stimuli:

Participants viewed images independently on separate Gateway E-3400 Pentium III 1000 Mega Hertz PC computers. All computers were equipped with Windows 2000 Professional Edition, 256MB of memory, and ATI RAGE 128 PRO Ultra GL AGP video cards.

These computers had 17 inch monitors with a resolution of 1024×768 pixels and display of 16 million colors. Software written in Java by the second author displayed the images to participants. (Lucas et al., 2005, p. 45)

If your equipment was custom made, provide a more detailed description. In the case of complex equipment, you may want to include a diagram. If you report any physical measurements, such as height, weight, or distance, you should report them in metric units. The *PM* contains extensive sections describing the use of metric measurements and their correct abbreviations (see pp. 130–136).

If your “equipment” consists of standardized psychological testing materials, then the label **testing instrument(s) subsection** (or measures) would be appropriate. For example, Niki James (a student at the University of Evansville, Evansville, Indiana) and Mary Pritchard (a faculty member at Boise State University, Boise, Idaho) described the instrument they used to measure stress in college students in the following manner:

The Inventory of College Students Recent Life Experiences (Kohn, Lafreniere, & Gurevich, 1990) was used to assess stressful events particularly related to college students’ lives (e.g., “Dissatisfaction with school”). Participants were asked to rate how much these responses have been a part of their lives in the past month on a 4-point scale (1 = *not at all a part of my life*, 4 = *very much part of my life*). This is a highly reliable and valid measure (Kohn et al., 1990). (James & Pritchard, 2005, p. 63)

If your experiment entailed the use of more than one category of materials, apparatus, or testing instruments, you should combine the relevant names when labeling this subsection.

The *PM* contains a short discussion about this subsection on pages 19–20.

Procedure subsection

The third subsection of the method section. Provides a step-by-step account of what the participants and experimenter did during the experiment.

Procedure The **procedure subsection** (see Figures 14-5 and 14-6) summarizes how you conducted the experiment. In addition to describing the steps that you followed, you must include a description of the experimental manipulations and controls (see Chapter 6), such as randomization, balancing, constancy, or counterbalancing, that you employed. Summarize any instructions you used unless they are unusual or complex. In the latter instance, you may want to present the instructions word for word.

If the experiment is involved and has several stages or phases, the procedure section can become lengthy. Dickson’s procedure section is intermediate in length. In contrast, the procedure involved in the administration of a questionnaire may be straightforward and brief. For example, in their study of drawing self-efficacy, Dunham and Yandell (2005) indicated the following:

The participants first signed a consent form. They were then given the Dunham Drawing Self-Efficacy Scale. There were two sessions of non-art students ranging between 2 and 14 participants,

there was one session of beginning art students with 12 participants, and there were two sessions of advanced art students ranging between four and five participants. After each group completed the self-efficacy scales, the group was debriefed. Each session took approximately 15 minutes. (p. 18)

Your primary goal in the procedure subsection is to describe how you conducted your experiment. You should give enough information to allow a replication of your method, but do not include unnecessary details (e.g., note that you recorded times with a stopwatch—the brand and model number would be overkill).



What critical methodological details do you find in Dickson's procedure subsection?

Critical details are pertinent points that another experimenter would have to copy in a replication study. We believe that the following details are critical:

- 2 × 2 design
- between-subjects design
- recruited participants from various locations on campus
- order of presentation of materials
- random assignment
- DV questions embedded among filler questions
- debriefing after participation
- 4 experimental conditions (type of artist: Black rap, White rap, Black country, White country)

Did we include any details that you left out? If so, reread the procedure subsection to see why we think that detail is absolutely necessary. For example, did you ignore the method of participant recruitment? We believe that this detail clarifies our earlier question about whether participants received any inducement to take part in the experiment—it appears unlikely from this statement.

The procedure subsection is typically the longest of the three components of the method section. Its length will vary depending on the complexity of your experiment. Dickson's procedure section is shorter than the materials section, but her procedure was rather simple, and she had to include more detail about the materials used. To read more about the procedure subsection, check page 20 of the *PM*.

Finally, you should not be surprised to find variations in researchers' method sections. Sections with combined names (e.g., *Design and Procedure* in Dickson's case) or with subsections in different orders are not unusual (Dickson's *Materials* section came after *Procedure*). The key point is that experimenters should provide you with the information you need to understand what they did in their experiment and how they did it.

Results

Results section The third major section of the APA-format paper. Contains information about the statistical findings from the experiment.

We introduced the format for the **results section** in Chapter 10 and reinforced those ideas in Chapters 11 and 12 when we discussed the notion of translating statistics into words. We do not use the word *translating* lightly—to some people, statistics resembles a foreign language. It is your job in the results section to decode the meaning of your numbers into words for the reader. At the same time you must provide the factual, numerical basis to back your decoding. The PM covers the presentation of statistical results on pages 20–26 and 136–147. Figure 14-7 contains Dickson’s results section. As you conduct your library research, you may notice that sometimes authors combine the results and discussion sections into one section in a published journal article. This combination is common for certain journals or for shorter experimental articles. It is likely that your instructor will want you to keep your sections separate.

Inferential Statistics As you write the results section, you should assume that your reader has a good understanding of statistics; therefore, you do not review basic concepts such as how to reject the null hypothesis and so on. The most important information to report is the specific findings from your inferential statistics. In Dickson’s paper, you see an example of how to report results from factorial ANOVA (F tests; see Chapter 12 for review) and post hoc t tests.



Why did Dickson use a factorial ANOVA to analyze her data? What was or were her IV(s)? What was or were her DV(s)?

First, you should remember from Chapter 12 that a factorial ANOVA is appropriate when you have more than one IV—thus, you should look for more than one IV in Dickson’s experiment. One focus of the experiment was race of the performer, so one IV was race (Black vs. White). The other IV was the genre of the performer’s music (rap vs. country).

In the procedure and materials subsections, you found out that Dickson asked the participants to rate the performers on four scales and then averaged the scores on the four scales. The experiment, therefore, had one DV: the overall index rating of the performers.

In presenting inferential statistical results, you must present the test that was used, the degrees of freedom for the test, the test statistics, the probability level, and a measure of effect size. In looking at Dickson’s results, the second sentence provides us a good example:

Data analysis revealed a significant main effect of race such that Black performers ($M = 4.32$, $SD = .91$) were rated more positively than were White performers ($M = 3.76$, $SD = 1.00$), $F(1, 92) = 10.42$, $p = .002$, $\eta^2 = .10$.

Notice that the statistical findings at the end of the sentence give us all five pieces of information: an F test (analysis of variance) was used, there were 1 and 92 degrees of freedom, the calculated test value was 10.42, the probability of these data occurring if the null hypothesis is true was 2 in 1,000, and the effect size was small (.10). You should present the same type of information for findings relevant to the experimental hypothesis

Eminem Versus Charley 7

labeled at 1 (*not at all*) and 7 (*very much*): (a) "Overall, how much do you like this performer?"; (b) "How talented do you think this performer is?"; (c) "How legitimate is this performer?"; and finally (d) "How successful do you predict this performer will be in the music industry?" Because these items were highly correlated, they were combined into an overall index reflecting participants' favorability of the performer (Cronbach's $\alpha = .80$). The scores on the four items were averaged together, and consistent with the scales on the individual items, the overall index is on a 7-point scale with higher numbers representing a more favorable perception of the performer and lower numbers indicating an unfavorable perception.

Results

A 2 (race of performer: Black or White) \times 2 (genre of music: rap or country) between-subjects analysis of variance (ANOVA) was performed on the index assessing the favorability of the performer (i.e., likability, perceived talent, perceived legitimacy, and predicted success). Data analysis revealed a significant main effect of race such that Black performers ($M = 4.32$, $SD = .91$) were rated more positively than were White performers ($M = 3.76$, $SD = 1.00$), $F(1, 92) = 10.42$, $p = .002$, $\eta^2 = .10$. However, this main effect was qualified by the significant two-way interaction between race of performer and genre of music, which was consistent with predictions, $F(1, 92) = 26.72$, $p = .0001$, $\eta^2 = .23$. An inspection of the means in Figure 1 shows that participants rated a Black rap artist ($M = 4.79$, $SD = .68$) more favorably than a Black country artist ($M = 3.85$, $SD = .87$), $t(46) = 4.17$, $p = .0001$; however, they judged a White country artist ($M = 4.19$, $SD = .92$) more favorably than a White rap artist ($M = 3.33$, $SD = .91$), $t(46) = 3.24$, $p = .0001$. There was no main effect of genre of music, $F < 1$, ns. Participants reported similar ratings for rap performers ($M = 4.06$, $SD = 1.09$) and country performers ($M = 4.02$, $SD = .90$).

Discussion

The present study integrated and expanded on two lines of research. First, it extended previous research (Fried, 1996, 1999) on stereotyping involving the music industry. Whereas Fried (1996, 1999) examined reactions to music lyrics based on either the genre label or the performer's race, our study considered the interactive effect of both the race of the performer and the genre of music on people's evaluations of performers. The current study also differs from previous research conducted by Fried (1996, 1999) in that the lyrics in our study did not attempt to convey negative images such as violence or aggression. That is, Fried (1996, 1999) asked participants specifically

FIGURE 14-7 (M7). Results section; portion of discussion section.

Source: Adapted from "Relating Pretrial Publicity, Judicial Instruction, and Thought Suppression with Guilt Ratings," by V. M. Gauthraux, 2000, *Psi Chi Journal of Undergraduate Research*, 5, pp. 21–28. Copyright © 2000 Psi Chi, The National Honor Society in Psychology (www.psichi.org). Reprinted by permission. All rights reserved.

even if you do not find statistical significance. Although information from different test statistics may appear in a slightly different fashion, these five basic pieces of information are always necessary. We presented a set of *t* test results in Chapter 10 to which you can refer. You can find examples of how to present other statistical test results on pages 138–139 of the *PM*.

Descriptive Statistics In order to give a full picture of the data, it is customary to present descriptive statistics in addition to the inferential statistics. Means and standard deviations typically allow readers to get a good feel for the data. With a small number of groups, you can present the descriptive statistics in the text (as Dickson did), as we showed you in Chapter 10. On the other hand, with many groups, it may be more efficient and clearer to present the descriptive statistics in either a table or a figure, which we will discuss next.

Complementary Information In presenting your results, you must first decide how best to give the reader the necessary information. If your statistical information is relatively simple, merely reporting your findings in words and numbers is usually adequate. For more complex analyses or results, you may wish to include tables or figures to explicate your words and numbers.

Figure A pictorial representation of a set of results.

Figures Because Dickson's DV showed a significant interaction, she used a **figure** to clarify the presentation of her results. The figure appears in Figure 14-8. Figures can take the form of graphs (line, bar, circle, pie, scatter, or pictorial graphs), charts, dot maps, drawings, or photographs.

The *PM* presents information regarding figures on pages 176–201. As you can guess from the number of pages devoted to figures, they can be difficult to produce. The vast majority of the *PM* information about figures involves preparing figures for publication. Because you will probably be submitting your research paper as a class requirement, your instructor may not be a stickler for enforcing every APA requirement on figures you submit with your paper. For example, you can see from Figure 14-9 that figure caption(s) appear on a separate page in a manuscript prepared for publication. Your instructor may choose to have you put the caption on the page with each figure.

It is likely that the majority of figures you use will be line graphs or bar graphs (see Chapter 9) that depict findings similar to Dickson's. Graphs are particularly good for showing interaction patterns. Because we showed interaction patterns in Chapter 12 with line graphs, we have reformatted Dickson's table as a line graph in Figure 14-10 (see p. 352). It may be easier for you to see the interaction from the line graph than from the bar graph. Regardless of what type of figure you decide to use, you probably have access to a number of good software programs for graphing. For your class project, ask your instructor about his or her preference as to how you should produce a figure. Regardless of how you create your figures, one rule is constant: Be certain that you refer to your figures in the text at an appropriate place. This reference will cue the reader to look at the figure in order to process its meaning.

Table A chart containing an array of descriptive statistics.

Tables A **table** consists of a display of data, usually in numerical form. Tables are an alternative method of presenting data. To use a table, your data display should be large enough that it would be difficult or confusing to present it in the text.

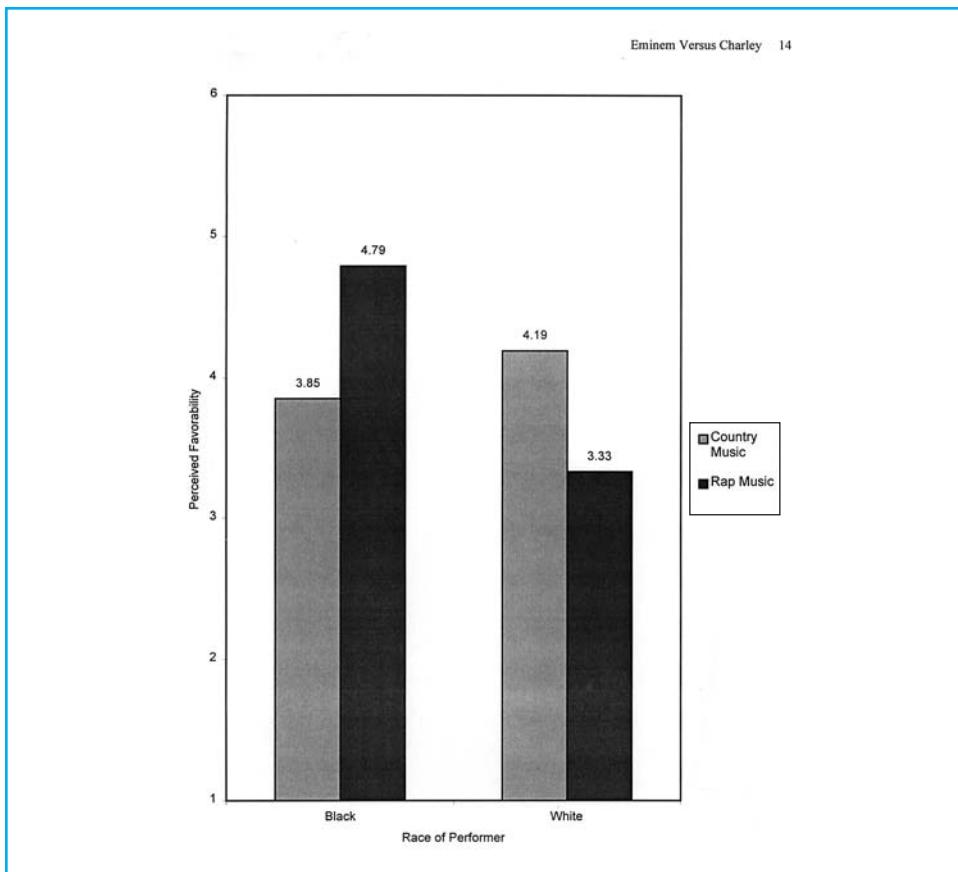


FIGURE 14-8 (M14). Dickson's figure.

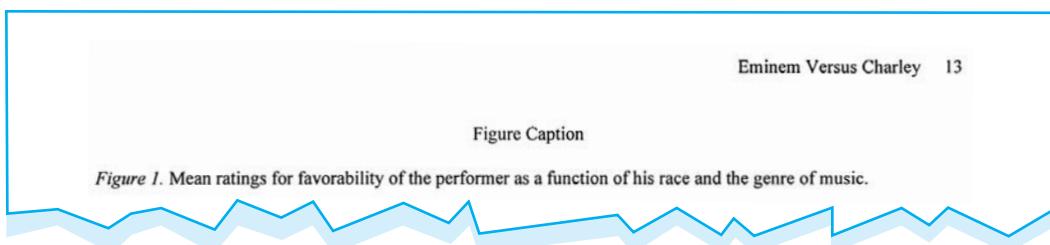
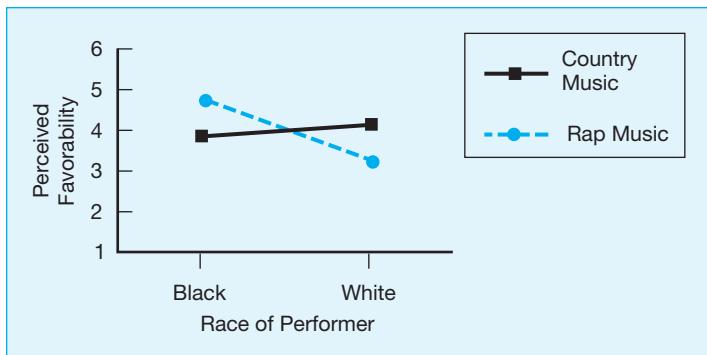


FIGURE 14-9 (M13). Caption for Dickson's figure.

**FIGURE 14-10**

Adaption of Dickson's results to line graph.

Table 1

Perceived Favorability of Performance Based on Performer's Race and Music Genre

Performer's Race		
Music Genre	Black	White
Rap		
<i>M</i>	3.85	4.19
<i>SD</i>	.87	.92
Country		
<i>M</i>	4.79	3.33
<i>SD</i>	.68	.91

Note. Favorability ratings made on a scale from 1 (negative) to 7 (positive).

FIGURE 14-11 Adaption of Dickson's results to table.

You often have to decide between using a table or a figure to present your data. Notice that Dickson presented her data only with a figure, not a table. We have adapted the data from her Figure 1 into the format that you see in Figure 14-11 (Dickson would have included a manuscript page header and page number if this page were actually in her manuscript).

The advantage of the table over the figure is that standard deviations can be included in the table. The advantage of the figure over the table is accessibility; the data, particularly the significant interaction, seem easier to understand and conceptualize in the pictorial presentation. Considering these advantages, the figure seemed a better choice in this situation. For your experimental report, your decision may be different. APA presents many guidelines for developing tables (see pp. 147–176 of the *PM*). You should check with your instructor to find out which guidelines you must follow in your paper. Because you will typically generate a table directly in your word-processing program, you may find a table easier to produce than a figure. You should, however, make your choice of a table versus a figure based on the quality of information provided to the reader rather than on ease of production. Again, if you use a table in your paper, be sure to refer to it at the appropriate point.



Can you think of a specific situation in which you would prefer to use a table instead of a figure?

The most obvious answer to this question is a situation in which you have a large number of means to present but for which there is not a significant interaction. We hope that a comparison of Figures 14-8 and 14-11 convinced you of the benefit of a figure for portraying interactions.

Discussion

You can find Dickson's **discussion section** in Figures 14-7, 14-12, and 14-13. According to the *PM* (pp. 26–27), you can address such issues as problem choice, levels of analysis, and application and synthesis in the discussion section. We like the somewhat simpler and clearer guidelines laid out in the previous edition of the *PM* (APA, 1994, p. 19), which recommended that you be guided by three questions in the discussion section:

- *What have I contributed here?*
- *How has my study helped to resolve the original problem?*
- *What conclusions and theoretical implications can I draw from my study?*

Typically, authors answer these three questions by (a) briefly restating their findings, (b) comparing and contrasting their findings to previous research cited in the introduction, and (c) giving their interpretation of their findings.

Discussion section The fourth major section of the APA-format paper. Contains a summary of the experiment's results, a comparison of those results to previous research, and the conclusion(s) drawn from the experiment.

Restating Results Your first task in the discussion section is to recap your results as briefly as possible. Typically, you will summarize only your significant findings, unless a null finding is particularly meaningful. If you conducted a large study with many outcomes, you may wish to feature only the most important findings at this point—usually those findings that have some bearing on your experimental hypothesis. This summary ensures that the reader can extract the most important information from your results section.

If you examine Dickson's (2001) discussion section, you will find that she used the three techniques listed just before this section. In the first paragraph, she summarized the information from the results section in one sentence: "In our study, participants judged Black rap artists more favorably than Black country music performers, whereas they judged White country music performers more favorably than White rap artists" (Dickson et al., 2001, p. 178). The sentence describes the significant interaction that Dickson found rather than the main effect of race because the interaction qualified the race effect.

Comparing Results to Previous Research It is important for you to evaluate how your results "stack up" against previous findings in your area of research that you summarized in your introduction. The previous research studies will be related to, but not the same as, your experiment. Typically, you will have made a prediction before the experiment about your expected findings based on your survey of previous research. You should tell the reader how

about their negative impressions of the song lyrics (e.g., its offensiveness, its threat to society, its need for warning labels), whereas the current study emphasized overall impressions of the performer. Rap may be more often associated with negative topics, but these topics can be found in both rap and country music. We chose to use neutral lyrics in the present study because we were not focusing on the relation between negative themes and music genre. Instead, the current study was based on findings that rap is associated with a Black culture that has both positive and negative attributes (Brigham, 1971). We examined the connection between music genre and the stereotype-consistent or stereotype-inconsistent race of the performer. In our study, participants judged Black rap artists more favorably than Black country music performers, whereas they judged White country music performers more favorably than White rap artists. These results support the hypothesis that individuals who deviate from societal expectations are judged more negatively than are individuals who adhere to social norms (Jackson et al., 1993; Knight et al., 2001). Manis, Nelson, and Shedler (1988) found that extreme stereotypes yielded contrast effects when behavior was discrepant from the established stereotype. In the present study, Black country music performers and White rap artists contrast from fairly ingrained societal expectations and thus received more negative judgments than the performers who adhere to societal norms.

The present study could be extended in a number of ways. For example, like Fried's (1996, 1999) research, the current study also presented the lyric sample to participants on paper. It would be interesting to explore whether an audio-recorded lyric sample would affect participants' evaluations of performers or the music. Perhaps auditory processing and visual processing of stereotype-consistent and stereotype-inconsistent information differ. In addition, the present study could broaden its scope by including a more varied sample of participants. That is, future studies could incorporate participants of different races and ages. The data in the current study were collected primarily from White undergraduate students at a liberal arts university, and the results from such a homogenous sample may not necessarily generalize to alternative populations or settings. Furthermore, exploring alternate stereotype violations could support the findings of the present study, and one area of interest could be the world of sports. For instance, Black hockey players violate societal expectations similar to the apparent violation made by Black country music performers and White rap performers. According to our findings, Black hockey players would receive more negative evaluations than would White hockey players because their behavior is inconsistent with societal expectations.

FIGURE 14-12 (M8). Second page of discussion.

Eminem Versus Charley 9

Because the population of other minority groups (e.g., Latinos) is approaching that of the Black population in America, future research concerning racial stereotypes in the music industry could examine the impact of the increasing popularity of Latin music (Gonzales, 1990). Garcia and Zapatel (2000) recently examined how the labels *Black rap*, *Latino rap*, and *alternative music* influence perceptions made by both Anglo and Hispanic participants and found that participants' perceptions of music differed depending on their own race. Specifically, Hispanic participants judged music labeled *Latino rap* more positively than music labeled *alternative*, whereas Anglo participants rated music labeled *alternative* more favorably than music labeled *Latino rap*. Similar to previous findings (Jackson et al., 1993), out-group categories (i.e., *Latino rap* for Anglo participants and *alternative* music for Hispanic participants) were judged more negatively than were categories that corresponded to the participants' ethnicity. Extended to a more diverse sample, the present study could offer support for the idea that out-group categories are judged more negatively than in-group categories by examining the relation between participants' ethnicities and their evaluations of rap and country music. Thus, regardless of the race of the performer, Black participants would be expected to rate country music more negatively than rap music, whereas White participants would be expected to judge rap music more negatively than country music. Furthermore, it would be interesting to determine whether or not performers of Latin music will remain primarily Latin and to examine perceptions of stereotype-consistent and stereotype-inconsistent performers in this genre. If future performers do deviate from racial stereotypes, the present study suggests that non-Latino performers of Latin music would be perceived less favorably than would Latino performers.

In closing, Eminem and Charley Pride share a connection in that their actions deviate from widespread stereotypes that specific races are associated with certain types of music. Although people may recognize these two performers because of their musical talent, it is more likely that they are recognized because they were bold enough to defy stereotypes in the music industry. By being a White rap artist and a Black country music performer, Eminem and Charley Pride became forerunners for performers who do not adhere to social norms, and they may have possibly influenced numerous music fans to expect the unexpected.

FIGURE 14-13 (M9). Final page of discussion.

accurate your predictions were. This information will help the reader to draw conclusions. For example, if you correctly predicted your results from previous research, the previous research and your study are both validated. On the other hand, if your prediction is not borne out, some doubt is cast—either your research or the previous research may be flawed in some way.

Looking at the first paragraph of Dickson's discussion section, you will see the comparison to previous research findings. In this case the previous research studies were related to Dickson's study because they dealt with stereotyping and music. On the other hand, these studies were different because they did not specifically combine performer's race and genre of music. Dickson's study demonstrated an interaction between those two variables, which was an important finding.

Interpreting the Results This portion of the discussion section gives you more of a free hand to engage in conjecture and speculation than any other portion of the experimental write-up. It is here that you draw the "bottom line" to your study: What is your overall conclusion? What are the implications of your results for any psychological theories? How can your results be applied in various settings—the laboratory, the real world, our body of psychological knowledge? What new research should grow out of this study? As you can see, there are a variety of questions you can address in your discussion section. Not all of these questions are appropriate for every study—pick the ones that are most important for your particular experiment.

The interpretation of Dickson's results appears in the last three sentences of the discussion section's first paragraph. The results of this study supported the notion that people who deviate from societal expectations are judged more harshly than people who conform to stereotypes.

The second and third paragraphs of Dickson's discussion section include possible directions for future research.



How many ideas for further research can you find in these two paragraphs?

We see four possible experiments mentioned:

1. Using audio-recorded lyrics rather than printed lyrics
2. Using participants from populations other than White students
3. Exploring racial stereotype expectations in sports settings
4. Extending the music genre to include Latin music

You may wish to mentally file this example away for future reference. As you read discussion sections of published articles, you might find ideas for your own research projects.

References

There are two important reasons why it is your responsibility to provide a complete list of accurate references to any published works cited in your research report. First, you must give credit to the original author(s) for any ideas and information that you got from reading their works. If you take exact wordings, paraphrases, or even ideas from an author without giving credit for that source, you are guilty of **plagiarism** (see Chapter 2). Second, you are providing a historical account of the sources you used in the event that a reader might wish to read them in their entirety. Have you

Plagiarism Using someone else's work without giving credit to the original source.

ever worked on a term paper and “lost” a good source that you had read because you didn’t write down all the necessary bibliographic information and couldn’t find it again? Most of us have had this experience, and it is quite frustrating. You can prevent that frustrating experience from happening by providing your reader with a complete and accurate reference list. The *PM*, which describes the reference list on page 28, provides general pointers about the list as well as examples of APA format for 95 different types of sources on pages 239–281. Reference formatting is so important that the *PM* now devotes a separate chapter (Chapter 4, pp. 215–281) to this topic.

Before proceeding, we must distinguish between a reference list and a bibliography, which you might have learned about in English classes. The reference list is *not* a list of every source that you read when you were writing your introduction and planning your experiment. The only references that you list are those from which you actually obtained information and that you cited somewhere in your paper. If you do not cite a particular source, you should not reference it.

The list of references begins on a new page after the end of your discussion section. You will find Dickson’s references in Figures 14-14 and 14-15.



Look at Figures 14-14 and 14-15. How many different types of references can you find? Can you identify each type?

Five different types of references appear in Dickson’s list:

1. *Book*: Brehm, Kassin, & Fein; Fiske & Taylor; Gonzales
2. *Journal article*: Brigham; Devine; Dovidio, Evans, & Tyler; Fried (both); Gaertner & McLaughlin; Jackson, Sullivan, & Hodge; Knight, Giuliano, & Sanchez-Ross; Manis, Nelson, & Shedler; Sagar & Schofield
3. *Chapter in edited book*: Dovidio, Brigham, Johnson, & Gaertner; Dovidio & Gaertner; Gaertner & Dovidio
4. *Conference presentation*: Garcia & Zapatael
5. *Music recording*: May; May & Staffell

As you look at a reference, you will find that the information about author(s) and date is listed first. This location makes it easy for the reader to see an author and date citation in your text and then to find the corresponding reference in your reference list. The reference list is alphabetized by the surname of the first author. If you have more than one article by the same first author, you alphabetize by the name of the second author. If the author information for two or more articles is identical, you should arrange the references by date, with the earliest article listed first. If the author information and dates are identical, alphabetize by the first main word of the title, and add lowercase letters (a, b, etc.) to the date to differentiate between the articles, as we previously mentioned.

The title of the scholarly work is the next piece of information, followed by supplementary information that helps a reader locate that work. As you can see, the supplementary

References

- Brehm, S. S., Kassin, S. M., & Fein, S. (1999). *Social psychology* (4th ed.). Boston: Houghton Mifflin.
- Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin*, 76, 15-38.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.
- Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice, and discrimination: Another look. In N. C. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes & stereotyping* (pp. 276-322). New York: Guilford Press.
- Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, 22, 22-37.
- Dovidio, J. F., & Gaertner, S. L. (1991). Changes in the expression and assessment of racial prejudice. In H. R. Knopke, R. J. Norrell, & R. W. Rogers (Eds.), *Opening doors: Perspectives on race relations in contemporary America* (pp. 119-148). Tuscaloosa: University of Alabama Press.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- Fried, C. B. (1996). Bad rap for rap: Bias in reactions to music lyrics. *Journal of Applied Social Psychology*, 26, 2135-2146.
- Fried, C. B. (1999). Who's afraid of rap: Differential reactions to music lyrics. *Journal of Applied Social Psychology*, 29, 705-721.
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism: Theory and research* (pp. 61-89). Orlando, FL: Academic Press.
- Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly*, 46, 23-30.
- Garcia, S. D., & Zapatel, J. P. (2000, January). *Perceptions of lyrics as a function of ethnicity and music genre*. Paper presented at the annual meeting of the Social Psychologists in Texas, San Antonio.
- Gonzales, J. L., Jr. (1990). *Racial and ethnic groups in America: A collection of readings*. Dubuque, IA: Kendall/Hunt.

FIGURE 14-14 (M10). First page of reference list.

- Jackson, L. A., Sullivan, L. A., & Hodge, C. N. (1993). Stereotype effects on attributions, predictions, and evaluations: No two social judgments are quite alike. *Journal of Personality and Social Psychology*, 65, 69-84.
- Knight, J. L., Giuliano, T. A., & Sanchez-Ross, M. G. (2001). Famous or infamous? The influence of celebrity status and race on perceptions of responsibility for rape. *Basic and Applied Social Psychology*, 23, 183-190.
- Manis, M., Nelson, T. E., & Shedler, J. (1988). Stereotypes and social judgment: Extremity, assimilation, and contrast. *Journal of Personality and Social Psychology*, 55, 28-36.
- May, B. (1973). The night comes down. [Recorded by Queen]. On *Queen* [Record]. London: EMI Records.
- May, B., & Staffell, T. (1973). Doing all right. [Recorded by Queen]. On *Queen* [Record]. London: EMI Records.
- Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of Personality & Social Psychology*, 39, 590-598.



FIGURE 14-15 (M11). Last page of reference list.

information differs depending on the particular type of reference you are using. Let's take a look at the general format for the three different types of references you are most likely to use in your papers: periodical articles, books, and chapters from edited books.

Periodical Articles The *PM* shows examples of 22 different types of references to periodicals on pages 240–247. Your most typical use of periodicals will probably be to reference articles in journals. The general format for periodicals is as follows (example adapted from APA, 2001, p. 239):

Author, A. A., Author, B. B., & Author, C. C. (date). Title of article. *Title of Periodical*, vol, ppp–ppp.

To see some examples, you can examine the 10 references shown in Figures 14-14 and 14-15, which we discussed in the answer to the last *Psychological Detective* question. The last names and initials of *all* authors appear in the same order as they appear in the journal article. You will remember, earlier in this chapter, we mentioned using et al. for multiple-author works. You typically *do not* use et al. in the reference list. List all authors except when there are more than six authors—then you list the first six in typical reference format and add et al. after the sixth name. The date refers to the year in which the particular journal issue containing the article was published. Be careful here—you may have to look at the front of the journal to get this information.

Type the title of the article with *only* the first word capitalized and *not* italicized. If the article title includes a colon, capitalize the first word after the colon also. In addition, you should capitalize any words that are normally capitalized (e.g., names, state, organizations, test names).

On the other hand, type the journal title with all primary words capitalized (APA's definition is any word of four letters or more). Words such as *a*, *and*, and *the* are not capitalized unless they are the first word or follow a colon. The journal title *is* italicized, as is the volume number of the journal (and its following comma) that immediately follows the journal title. Type only the volume number—do not precede it by *Vol.* or *V.* Volumes of journals sometimes also have issue numbers. For example, volume 59 represents the 2004 issues of *American Psychologist*; each month's issue is represented by its own number. Thus, the January 2004 issue would be Volume 59, Number 1, and so on. Most journals use continuous pagination throughout a volume; that is, the first issue of a new volume begins on page 1 and the pages run continuously until the next volume, which begins with page 1. In this case the issue number is not needed to find the referenced article and does not appear in the reference. A few journals, however, repaginate; that is, they begin *each issue* with page 1. If this is the case, then the issue number is necessary to find the article and is included in the reference. This reference format follows (example adapted from APA, 2001, p. 240):

Author, A. A., Author, B. B., & Author, C. C. (date). Title of article. *Title of Periodical*, vol(n),
ppp–ppp.

Notice that the issue number (n) appears within parentheses (no space after the volume number) and is *not* italicized. Finally, give the inclusive page numbers of the article—numbers only, no *p.* or *pp.* preceding.

Books The *PM* provides 11 examples of references to books on pages 248–251. This category consists of references to entire books rather than to chapters in edited books. The general format for book references is as follows (example adapted from APA, 2001, p. 248):

Author, A. A. (date). *Title of work*. Location: Publisher.

We have used a one-author example here, but don't be misled by it. You should include all authors' names and initials (see Brehm, Kassin, & Fein in Dickson's reference list), just as in the journal examples given previously. Use the date of the original publication of the book, which usually appears at the front of the book facing the title page.

The book title's format is a combination of the styles seen earlier for article titles and journal titles. It follows the style for an article title in terms of capitalization—you capitalize only the first word (and the first word after a colon or normally capitalized words)—however, like a journal title, the book's title is italicized.

The location and the name of the publisher appear in the last portion of the reference. If the city, such as New York, is well known for publishing, you type it alone (see p. 217 of the *PM* for a complete list); otherwise, you must give the city and state (two-letter postal abbreviation) or country. Many publishers now have offices in several locations—typically the first location listed is the one referenced. The name of the publisher is given in a brief form, omitting "superfluous terms, such as *Publishers, Co.*, or *Inc.*, which are not required to identify the publisher" (APA, 2001, p. 230). Many times you will find that a corporate author and publisher are the same, as with the *PM*. In such a case, rather than repeat the information for the publisher, you simply type "Author" after the location (e.g., see the reference to the *PM* in the reference list for this book). Dickson's manuscript has three examples of references to books.

Chapters From Edited Books Most edited books contain chapters written by different authors. This type of reference allows you to cite information from a single chapter within

such a book. The *PM* gives seven examples of such references on pages 252–255. The general format of such a reference is as follows (example adapted from APA, 2001, p. 239):

Author, A. A., & Author, B. B. (date). Title of chapter. In C. C. Editor, D. D. Editor, & E. E. Editor (Eds.), *Title of book* (pp. nnn–nnn). Location: Publisher.

As you can see, this type of reference is much like a journal article reference combined with a book reference. This example includes two authors and three editors, but any number of either is possible. You would provide the information concerning author(s) and date as we have previously discussed. The title of the chapter refers to the specific smaller work within the larger book. Capitalize the chapter title in the same manner as a journal article title: Capitalize only the first word, the first word after a colon, and words that are normally capitalized.

List all the editors' names. Notice that you do not reverse the initials and surnames. After a comma, you list the book's title (*italicized*), with capitalization in the same fashion as mentioned previously for a book title. The inclusive pages of the chapter appear parenthetically after the book's title to make it easier for the reader to locate the specific chapter in the book (be sure to type "pp."). Finally, include the location and publisher information as for any book. Dickson's manuscript includes three examples of chapters in edited books, as listed previously.

World Wide Web Sources Providing references for information obtained from the World Wide Web is just as important as providing references for written material—in fact, it may be even more important. Unlike journal articles or books, material on the Web can change quite rapidly. Authors can modify Web information from day to day, which can create a problem in keeping current. The *PM* has an extensive section dealing with electronic media references, which includes 25 examples of Internet and Web references, on pages 268–281. Because there are so many different types of electronic media sources, it is impossible to cover them all. We will provide an example of a citation to information on a Web site; if you have a different type of source, be sure to consult the *PM*.

American Psychological Association. (2003). *APA Style.org: Electronic references*. Retrieved July 17, 2005, from <http://www.apastyle.org/electref.html>

Note that this reference style is similar to a book reference; however, you include the actual retrieval site and date. This additional information allows a reader to find the site and to determine whether the information still exists (or perhaps has been updated) at that site. Providing an exact date of retrieval is extremely important for these reasons. Interestingly, APA style for electronic references is evolving quickly. You can go to the Web site listed in the reference earlier in this section for the latest updates or to help answer questions you may have.

The relevant text citation for this reference would be the same for a book; in this case it would be (American Psychological Association, 2003). If you wish to cite an entire Web site (rather than a specific document on the Web site), you can simply list the site in the text—no reference is necessary (e.g., <http://www.apa.org> for the APA site).

Other References Although we expect that most of your references will be to periodicals, books, and chapters in edited books, the *PM* has 30 examples of other types of references you might use. These other references include technical and research reports, proceedings of meetings and symposia, doctoral dissertations and master's theses, unpublished works and publications of limited circulation, reviews, and audiovisual media. No matter what type of material you wish to reference, the *PM* has a format for you.

A Disclaimer The fourth edition of the *PM* (APA, 1994) used a different reference format. Although all the information included was the same, the appearance of the references was different. In the older format, the initial line of each reference was indented and subsequent lines began at the left margin; however, when references were printed in journal articles, this margin situation was reversed—the first line began at the left margin and subsequent lines were indented (referred to as a “hanging indent”). With this type of formatting, the first authors’ names stand out and are easier to locate. Under the new (2001) guidelines, manuscript references are formatted exactly as they appear in publication.

We give you this explanation to alert you to the possibility that you might see an older student paper with references formatted in the old manner. It is not unusual for students to find previous copies of research papers to use as examples. If you copy that style of reference format, you will be in error. While examining the manuscript pages of Dickson’s paper and the published product, you may have noted other differences.



Can you figure out why an APA-formatted manuscript and a published article (or final copy) might differ in appearance?

APA format is intended to help authors and journal editors produce a product more easily. In a published article or final copy; however, the appearance of the document is the more important goal. It is difficult for instructors to decide whether to require strict APA format or to accept a more aesthetically appealing product.

Appendix

Most published articles do not contain an appendix because of space limitations; appendixes are more common in student papers (*PM*, 2001, p. 324). Typically you include information that might help readers understand some detail(s) of your study that would be distracting in the body of the paper. The *PM* (p. 28) gives as examples a new computer program, an unpublished test, a mathematical proof, a list of stimuli used, and the description of a complex piece of equipment.

Dickson included an appendix that contained the lyrics to the music used in her experiment. We have *not* included the appendix in the text because of space limitations (but the reference to the appendix appears in the “Materials” subsection in Figure 14-19).

Author Note

Author note A note at the end of an experimental report that contains information about the author or paper for readers’ use.

Your manuscript may have an **author note**, depending on your instructor’s preferences. If you do use an author note, begin it on a new page. As you can see from Dickson’s author note in Figure 14-16, she acknowledged several people who helped on the project and provided a name and address for readers to contact her for information or copies of the paper. Other information might be an acknowledgment of an earlier presentation of the paper (e.g., at a conference) or sources of

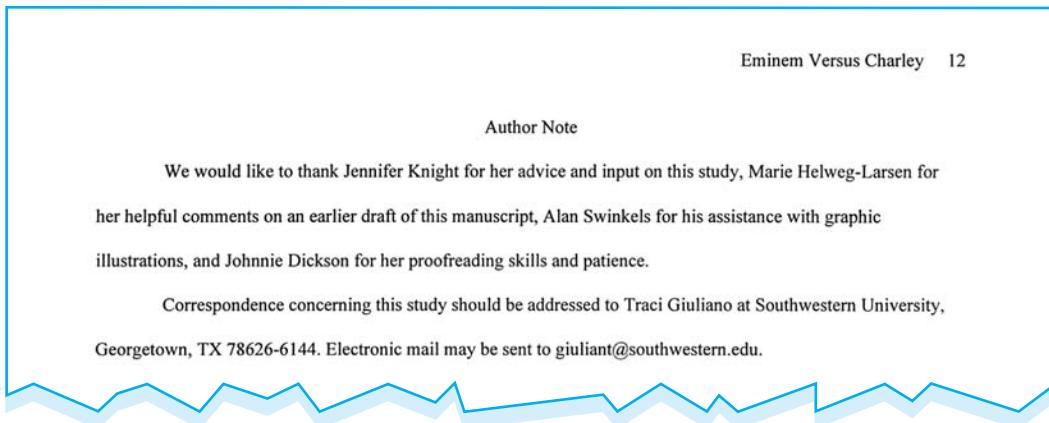


FIGURE 14-16 (M12). Author note page.

financial support making the study possible. The *PM* provides information about the author note on pages 29, 203–205, and 300; Dickson did not include all four paragraphs described on pages 203–205 because the journal did not require them. Your instructor may have you deviate from the *PM*'s standard format for a class report.

Headings

To recapitulate, APA-format papers use a different type of heading for each section of the report. The major sections of the report, such as the introduction, method, results, and discussion, are introduced by a Level 1 heading. Subsections within these main sections are introduced by lower level headings. For example, the participants, apparatus, and procedure subsections of the method section are generally introduced by a Level 3 heading. Level 3 headings are left-margin justified, are italicized, have the first letter of each major word capitalized, and occupy a line by themselves. Should you have to further subdivide these subsections, use a **Level 4 heading**. Level 4 headings are indented five spaces and italicized, have only the first letter of the first word capitalized, and end with a period. You begin typing on the same line following the period that concludes a Level 4 heading. Level 2 headings are rarely used except in multiexperiment studies; they are centered, italicized, and have capitalized main words. In a multiexperiment study, typical Level 1 headings become Level 2 (e.g., *Method*).

Table 14-1 summarizes the five types of section headings used in APA-format articles. Level 1 and Level 3 headings are the ones most frequently used in preparing a research report that describes a single experiment (as in Dickson's example). As we have seen, however, the description of a single experiment may require the use of Level 4 headings when the participants, apparatus, and procedure subsections are further subdivided (see Table 14-1). Likewise, when you are presenting more than one experiment, you will use Level 1, 2, 3, and 4 headings as shown in Table 14-1.

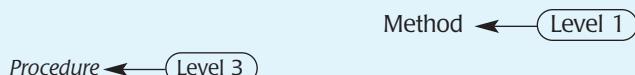
Level 4 heading A section title that is indented five spaces, italicized, has only the first word capitalized, and ends with a period; it does not occupy a separate line.

TABLE 14-1 Levels and Locations of Headings Used in an APA-Format Paper: Selecting the Levels of Heading

Find the section of your paper that breaks into the finest level of subordinate categories. Then use the guidelines that follow to determine the level, position, and arrangement of headings. Few articles require all levels of heading. Note that each subheading must have at least one counterpart at the same level within a section; for brevity, the examples that follow do not include counterparts.

One level. For a short article, one level of heading may be sufficient. In such cases, use only centered uppercase and lowercase headings (Level 1).

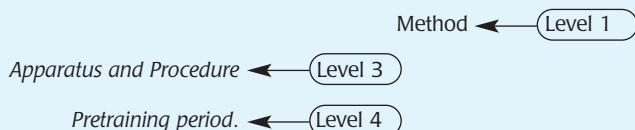
Two levels. For many articles in APA journals, two levels of headings meet the requirements. Use Level 1 and Level 3 headings:



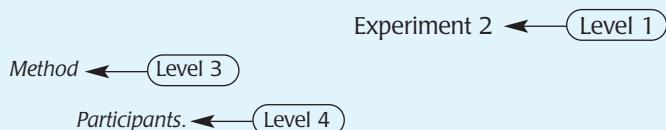
If the material subordinate to the Level 1 headings is short or if many Level 3 headings are necessary, indented, italicized lowercase paragraph headings (Level 4) may be more appropriate than Level 3 headings. (A Level 4 heading should apply to all text between it and the next heading, regardless of the heading level of the next heading.)

Three levels. For many articles, three levels of headings are needed. Use Level 1, Level 3, and Level 4 headings.

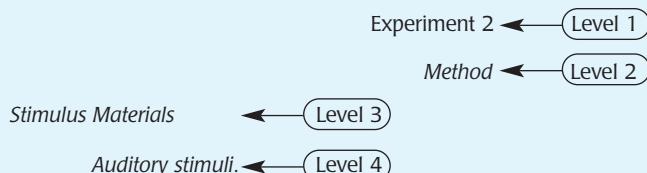
In a single-experiment study, these three levels of headings may look like this:



In a *multiexperiment study*, these three levels of headings may look like this:



Four levels. For some articles, particularly multiexperiment studies, monographs, and lengthy literature reviews, four levels of headings are needed. Use heading Levels 1 through 4:



Five levels. Occasionally, an article requires five levels of headings. In such cases, subordinate all four levels above by introducing a Level 5 heading—a centered uppercase heading—above the other four.

Source: This material has been reproduced from the *Publication Manual of the American Psychological Association*, 2001, pp. 114–115. Reprinted with permission. Neither the original nor this reproduction can be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without prior written permission of the American Psychological Association.

Eminem Versus Charley Pride: Race, Stereotypes, and Perceptions of Rap and Country Music Performers

ALLISON J. DICKSON

TRACI A. GIULIANO*

JAMES C. MORRIS

KERI L. CASS

Southwestern University

The present study explored the effects of stereotype deviation in the music industry on people's perceptions of performers. One hundred college students (48 men, 52 women) examined a profile of a fictitious performer containing a picture, a brief biography, and a lyric sample. As part of a 2-way between-subjects design, participants made judgments about either a Black or a White musician who performed either rap or country music. The results showed that a Black rap performer was rated more favorably than a Black country performer, and a White country performer was rated more favorably than a White rap performer. Consistent with predictions, people who violate societal expectations are judged more harshly than are people who conform to societal expectations, particularly in cases involving strong preexisting racial stereotypes.

WHAT DO EMINEM AND CHARLEY PRIDE HAVE in common? Perhaps the connection these two performers share is subtle, but they are in fact quite similar in at least one aspect of their careers. Both Charley Pride, a Black country music performer, and Eminem, a White rap performer, deviate from social expectations that are a part of the music industry. Specifically, these two musicians defy cultural stereotypes by performing types of music that are not typically associated with their race.

Racial stereotypes exist in most individuals, and they can influence subsequent judgments made by a perceiver (Devine, 1989; Dovidio, Evans, & Tyler, 1986; Gaertner & McLaughlin, 1983). For example, Gaertner and McLaughlin (1983) studied the effect of racial stereotypes on perceptions and found that White students responded faster to positive stereotyped words (e.g., *smart*) when the words followed the race *White* rather than *Black*. In addition, Sagar and Schofield (1980) examined the perceptions made by sixth-grade boys about ambiguous behavior. They found that both Black and White boys construed ambiguously aggressive behaviors (such as one child

bumping into another) as being more threatening if the actions were performed by a Black boy rather than a White boy. Most people today would not be likely to openly express racist beliefs, but the results of the above studies support the aversive racism perspective, which suggests that subtle and indirect forms of racism persist in society today (Dovidio & Gaertner, 1991; Gaertner & Dovidio, 1986). That is, although current cultural values emphasize fairness and racial equality, White individuals have a historic tradition of having negative beliefs concerning Blacks and other minority groups (Dovidio, Brigham, Johnson, & Gaertner, 1996). Consequently, racial stereotypes continue to exist and to influence interactions among individuals in society, but perhaps in more subtle ways.

Author note. We would like to thank Jennifer Knight for her advice and input on this study, Marie Helweg-Larsen for her helpful comments on an earlier draft of this manuscript, Alan Swinkels for his assistance with graphic illustrations, and Johnnie Dickson for her proofreading skills and patience.

Correspondence concerning this study should be addressed to Traci Giuliano at Southwestern University, Georgetown, TX 78626-6144. Electronic mail may be sent to giuliani@southwestern.edu.

PSI CHI JOURNAL OF UNDERGRADUATE RESEARCH □ Winter 2001

Copyright 2001 by Psi Chi, The National Honor Society in Psychology (Vol. 6, No. 4, 175–180 / ISSN 1089-4136). *Faculty Supervisor

175

FIGURE 14-17 (JA1).

Source: From "Eminem vs. Charley Pride: Race, Stereotypes and Perceptions of Rap and Country Music Performers," by A. J. Dickson, T. A. Giuliano, J. C. Morris, and K. L. Cass, 2001, *Psi Chi Journal of Undergraduate Research*, 6, pp. 175–180. Copyright © 2001 Psi Chi, The National Honor Society in Psychology (www.psichi.org). Reprinted by permission. All rights reserved.

RACE, STEREOTYPES, AND PERCEPTIONS OF PERFORMERS □ *Dickson, Giuliano, Morris, and Cass*

Because stereotypes can influence the judgments and behaviors of perceivers, deviations from a stereotype should have similar effects. In general, the expectations of the perceiver and the extent to which these expectations are confirmed or disconfirmed can influence judgments. When behavior only slightly varies from expectations, the difference might not be noticed, but perceivers often magnify the discrepancy when actions differ significantly from expectations. This phenomenon is known as the contrast effect (Brehm, Kassin, & Fein, 1999). In fact, a person who displays behavior inconsistent with societal expectations is often evaluated more extremely than is a person who behaves consistently with expectations (Knight, Giuliano, & Sanchez-Ross, 2001). Jackson, Sullivan, and Hodges (1993) examined the effects of describing stereotype-consistent or stereotype-inconsistent behavior of Black out-group targets and White in-group targets on social evaluations made by participants who assessed a college application. They found that people who deviate from a norm are judged more extremely than if they behave as the norm dictates. Specifically, stereotype-inconsistent Black applicants with strong credentials were evaluated more favorably than were strong White applicants, and stereotype-inconsistent White applicants with weak credentials were evaluated less favorably than were weak Black applicants.

Both stereotypes and stereotype-inconsistent behavior affect the evaluations individuals make about other people in a variety of social interactions. Fried (1996, 1999) studied biased reactions involving the music industry and found that individuals had very different reactions to music labeled as *rap* or *country* despite the fact that the song lyrics were exactly the same. In two studies, she found that people generally considered rap music to be more violent and more offensive than country music. Furthermore, a folk song that was presented as being performed by a Black artist was judged more negatively than the very same song when it was presented as being performed by a White artist (Fried, 1996). Fried (1999) attributed her results to stereotypes in that rap is usually associated with Black culture whereas country music is often thought of as being a part of White culture. By priming a Black stereotype with the use of the label *rap*, it is possible that individuals apply negative stereotypes that have been shown to be associated with African Americans (Brigham, 1971). Therefore, racial stereotypes can impact evaluations of music performers (Fried, 1996, 1999).

According to Fiske and Taylor (1991), long-held stereotypes are not easily altered, but modification of stereotypes may begin with a divergence by stereo-

typed individuals or groups. Previous research has examined the effects of stereotypes and stereotype deviation on people's evaluations of other individuals (Jackson et al., 1993; Knight et al., 2001). The present study attempted to integrate and expand on these concepts in relation to the music industry. Specifically, whereas Fried (1996, 1999) examined the effect of either the race of the performer or the labeled genre of music of a song on evaluations about the music itself, the design of the present experiment explored the interactive effects of the race of the performer and the genre of music on participants' evaluations of the performer. In doing so, we explored the difference between perceptions of persons who adhere to social expectations versus persons who deviate from the stereotype.

Consistent with previous research (Jackson et al., 1993; Knight et al., 2001), we expected that participants would judge performers who behave consistently with social norms (i.e., Black rap artists and White country performers) more favorably than performers who deviate from societal expectations (i.e., White rap artists and Black country performers). That is, because country music is associated with White culture, a Black country performer does not exhibit behavior consistent with this stereotype and, as a result, this performer should elicit negative judgments. The same reaction should occur with a White rap artist because his or her behavior is inconsistent with the stereotype that rap is predominantly a part of Black culture.

Method

Participants

Data were collected from 100 undergraduates (48 men, 52 women) at Southwestern University, a small liberal arts college in the Southwest. Demographically, the university is composed primarily of White, middle-to upper middle-class students; as such, the current sample (which was representative of the campus at large) consisted almost exclusively of White students. Participant volunteers ranged in age from 18 to 26 years ($M = 19.77$ years). Data from four participants were excluded from the analysis because these participants either failed to follow instructions or they did not pass the manipulation check. Specifically, these participants were unable to identify the race of the performer and genre of his music that was presented in their survey packet.

Design and Procedure

The present study used a 2 (race of performer: Black or White) \times 2 (genre of music: rap or country) between-subjects design to explore the effect of the

FIGURE 14-18 (JA2).

RACE, STEREOTYPES, AND PERCEPTIONS OF PERFORMERS □ Dickson, Giuliano, Morris, and Cass

race of a performer and the genre of music on perceptions of the performer. We recruited participants from various locations on campus and asked them to contribute to an investigation exploring "people's perceptions of music." Once they agreed, participants viewed a picture of a male performer, read a brief biography about him, and read a lyric sample of his music. Next, they completed a survey in which they made judgments about the performer and his music, and they responded to filler questions concerning their taste in music in general to corroborate the cover story. Each participant was randomly assigned to one of four experimental conditions and read a profile of either a Black rap artist, a White rap artist, a Black country performer, or a White country performer. Measures of the primary dependent variable (i.e., how favorably participants rated the performers) were embedded among filler questions in the survey. Following completion of the survey and a brief manipulation check that focused on the performer's race and the genre of music, we explained the purpose of the present study to the participants and asked them not to discuss it with anyone.

Materials

A three-page experimental packet, which ostensibly contained a survey about people's perceptions of music, was distributed along with an envelope. The first page contained a "performer profile" and included a color picture of a Black or White male performer, a brief biography about him indicating that he was either a rap or country performer, and a lyric sample from one of his songs. The subsequent pages contained the survey, which was used to measure participants' reactions to what they had seen and read on the previous page.

Each performer profile contained a biography that included the name of the performer (i.e., D.J. Jones), his hometown (Atlanta, Georgia), and a brief summary of his musical career (e.g., "D.J. has been singing since he was 14, and recently signed a record deal with a major country label. He will soon be on an international tour opening for a popular country artist"). To create the sample song lyrics, we slightly altered and combined two rock songs (May, 1973; May & Staffell, 1973; see Appendix A). With the exception of the two manipulations, we used the same biography and song lyrics in each of the four experimental conditions. The first manipulation altered the genre of music (i.e., rap or country) that the artist performed. Specifically, the name of the performer and the type of music he performed was labeled beneath his picture, and this label coincided with the genre of music described in the biography. Next, a

photograph manipulated the race of the performer so that each biography was accompanied by a picture of either a Black or a White man. To ensure that the only attribute differing between the Black and White performers was in fact race, we conducted a pilot test in a Research Methods class to match the Black and White performers on attractiveness and age. The participants in the pilot test were all White, which is representative of the ethnic composition of the participants in the present study. From a pool of 20 color photographs of nonfamous Black and White men selected from magazines, the two stimuli selected for the present study were most similar in perceived attractiveness and age.

Following the performer profile were the questions that assessed perceptions of the performer and his music, as well as the musical tastes of the participants. Embedded among demographic questions (e.g., sex and age) and other filler questions (e.g., "On average, how many CDs and/or tapes do you buy a month?") were the items that examined participants' perceptions of the performer. Specifically, participants rated on 7-point scales with endpoints labeled at 1 (*not at all*) and 7 (*very much*): (a) "Overall, how much do you like this performer?"; (b) "How talented do you think this performer is?"; (c) "How legitimate is this performer?"; and finally (d) "How successful do you predict this performer will be in the music industry?" Because these items were highly correlated, they were combined into an overall index reflecting participants' favorability of the performer (Cronbach's $\alpha = .80$). The scores on the four items were averaged together, and consistent with the scales on the individual items, the overall index is on a 7-point scale with higher numbers representing a more favorable perception of the performer and lower numbers indicating an unfavorable perception.

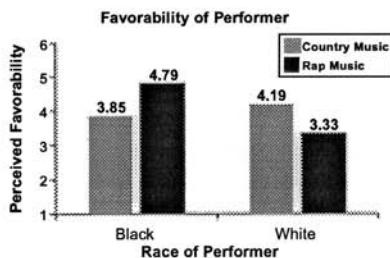
Results

A 2 (race of performer: Black or White) \times 2 (genre of music: rap or country) between-subjects analysis of variance (ANOVA) was performed on the index assessing the favorability of the performer (i.e., likability, perceived talent, perceived legitimacy, and predicted success). Data analysis revealed a significant main effect of race such that Black performers ($M = 4.32$, $SD = .91$) were rated more positively than were White performers ($M = 3.76$, $SD = 1.00$), $F(1, 92) = 10.42$, $p = .002$, $\eta^2 = .10$. However, this main effect was qualified by the significant two-way interaction between race of performer and genre of music, which was consistent with predictions, $F(1, 92) = 26.72$, $p = .0001$, $\eta^2 = .23$. An inspection of the means in Figure 1 shows that participants rated a Black rap artist ($M = 177$

RACE, STEREOTYPES, AND PERCEPTIONS OF PERFORMERS □ Dickson, Giuliano, Morris, and Cass

FIGURE 1

Mean ratings for favorability of the performer as a function of his race and the genre of music.



4.79, $SD = .68$) more favorably than a Black country artist ($M = 3.85$, $SD = .87$), $t(46) = 4.17$, $p = .0001$; however, they judged a White country artist ($M = 4.19$, $SD = .92$) more favorably than a White rap artist ($M = 3.33$, $SD = .91$), $t(46) = 3.24$, $p = .0001$. There was no main effect of genre of music, $F < 1$, ns : Participants reported similar ratings for rap performers ($M = 4.06$, $SD = 1.09$) and country performers ($M = 4.02$, $SD = .90$).

Discussion

The present study integrated and expanded on two lines of research. First, it extended previous research (Fried, 1996, 1999) on stereotyping involving the music industry. Whereas Fried (1996, 1999) examined reactions to music lyrics based on either the genre label or the performer's race, our study considered the interactive effect of both the race of the performer and the genre of music on people's evaluations of performers. The current study also differs from previous research conducted by Fried (1996, 1999) in that the lyrics in our study did not attempt to convey negative images such as violence or aggression. That is, Fried (1996, 1999) asked participants specifically about their negative impressions of the song lyrics (e.g., its offensiveness, its threat to society, its need for warning labels), whereas the current study emphasized overall impressions of the performer. Rap may be more often associated with negative topics, but these topics can be found in both rap and country music. We chose to use neutral lyrics in the present study because we were not focusing on the relation between negative themes and music genre. Instead, the current study was based on find-

ings that rap is associated with a Black culture that has both positive and negative attributes (Brigham, 1971). We examined the connection between music genre and the stereotype-consistent or stereotype-inconsistent race of the performer. In our study, participants judged Black rap artists more favorably than Black country music performers, whereas they judged White country music performers more favorably than White rap artists. These results support the hypothesis that individuals who deviate from societal expectations are judged more negatively than are individuals who adhere to social norms (Jackson et al., 1993; Knight et al., 2001). Manis, Nelson, and Shedler (1988) found that extreme stereotypes yielded contrast effects when behavior was discrepant from the established stereotype. In the present study, Black country music performers and White rap artists contrast from fairly ingrained societal expectations and thus received more negative judgments than the performers who adhere to societal norms.

The present study could be extended in a number of ways. For example, like Fried's (1996, 1999) research, the current study also presented the lyric sample to participants on paper. It would be interesting to explore whether an audio-recorded lyric sample would affect participants' evaluations of performers or the music. Perhaps auditory processing and visual processing of stereotype-consistent and stereotype-inconsistent information differ. In addition, the present study could broaden its scope by including a more varied sample of participants. That is, future studies could incorporate participants of different races and ages. The data in the current study were collected primarily from White undergraduate students at a liberal arts university, and the results from such a homogenous sample may not necessarily generalize to alternative populations or settings. Furthermore, exploring alternate stereotype violations could support the findings of the present study, and one area of interest could be the world of sports. For instance, Black hockey players violate societal expectations similar to the apparent violation made by Black country music performers and White rap performers. According to our findings, Black hockey players would receive more negative evaluations than would White hockey players because their behavior is inconsistent with societal expectations.

Because the population of other minority groups (e.g., Latinos) is approaching that of the Black population in America, future research concerning racial stereotypes in the music industry could examine the impact of the increasing popularity of Latin music (Gonzales, 1990). Garcia and Zapata (2000) recently examined how the labels *Black rap*, *Latino rap*, and

FIGURE 14-20 (JA4).

RACE, STEREOTYPES, AND PERCEPTIONS OF PERFORMERS □ Dickson, Giuliano, Morris, and Cass

alternative music influence perceptions made by both Anglo and Hispanic participants and found that participants' perceptions of music differed depending on their own race. Specifically, Hispanic participants judged music labeled *Latino rap* more positively than music labeled *alternative*, whereas Anglo participants rated music labeled *alternative* more favorably than music labeled *Latino rap*. Similar to previous findings (Jackson et al., 1993), out-group categories (i.e., *Latino rap* for Anglo participants and *alternative* music for Hispanic participants) were judged more negatively than were categories that corresponded to the participants' ethnicity. Extended to a more diverse sample, the present study could offer support for the idea that out-group categories are judged more negatively than in-group categories by examining the relation between participants' ethnicities and their evaluations of rap and country music. Thus, regardless of the race of the performer, Black participants would be expected to rate country music more negatively than rap music, whereas White participants would be expected to judge rap music more negatively than country music. Furthermore, it would be interesting to determine whether or not performers of Latin music will remain primarily Latin and to examine perceptions of stereotype-consistent and stereotype-inconsistent performers in this genre. If future performers do deviate from racial stereotypes, the present study suggests that non-Latino performers of Latin music would be perceived less favorably than would Latino performers.

In closing, Eminem and Charley Pride share a connection in that their actions deviate from widespread stereotypes that specific races are associated with certain types of music. Although people may recognize these two performers because of their musical talent, it is more likely that they are recognized because they were bold enough to defy stereotypes in the music industry. By being a White rap artist and a Black country music performer, Eminem and Charley Pride became forerunners for performers who do not adhere to social norms, and they may have possibly influenced numerous music fans to expect the unexpected.

References

- Brehm, S. S., Kassin, S. M., & Fein, S. (1999). *Social psychology* (4th ed.). Boston: Houghton Mifflin.
- Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin*, 76, 15–38.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18.
- Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice, and discrimination: Another look. In N. C. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotype & stereotyping* (pp. 278–322). New York: Guilford Press.
- Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, 22, 22–37.
- Dovidio, J. F., & Gaertner, S. L. (1991). Changes in the expression and assessment of racial prejudice. In H. R. Kropke, R. J. Norrell, & R. W. Rogers (Eds.), *Opening doors: Perspectives on race relations in contemporary America* (pp. 119–148). Tuscaloosa: University of Alabama Press.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- Fried, C. B. (1996). Bad rap for rap: Bias in reactions to music lyrics. *Journal of Applied Social Psychology*, 26, 2135–2146.
- Fried, C. B. (1999). Who's afraid of rap: Differential reactions to music lyrics. *Journal of Applied Social Psychology*, 29, 705–721.
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism: Theory and research* (pp. 61–89). Orlando, FL: Academic Press.
- Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly*, 46, 23–30.
- Garcia, S. D., & Zapata, J. P. (2000, January). *Perceptions of lyrics as a function of ethnicity and music genre*. Paper presented at the annual meeting of the Social Psychologists in Texas, San Antonio.
- Gonzales, J. L., Jr. (1990). *Racial and ethnic groups in America: A collection of readings*. Dubuque, IA: Kendall/Hunt.
- Jackson, L. A., Sullivan, L. A., & Hodge, C. N. (1993). Stereotype effects on attributions, predictions, and evaluations: No two social judgments are quite alike. *Journal of Personality and Social Psychology*, 65, 69–84.
- Knight, J. L., Giuliano, T. A., & Sanchez-Ross, M. G. (2001). Famous or infamous? The influence of celebrity status and race on perceptions of responsibility for rape. *Basic and Applied Social Psychology*, 23, 183–190.
- Manis, M., Nelson, T. E., & Shedler, J. (1988). Stereotypes and social judgment: Extremity, assimilation, and contrast. *Journal of Personality and Social Psychology*, 55, 28–36.
- May, B. (1973). The night comes down. [Recorded by Queen]. On *Queen* [Record]. London: EMI Records.
- May, B., & Staffell, T. (1973). Doing all right. [Recorded by Queen]. On *Queen* [Record]. London: EMI Records.
- Sager, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of Personality & Social Psychology*, 39, 590–598.

FIGURE 14-21 (JA5).

■ REVIEW SUMMARY

1. Psychologists have developed their own format, **APA style**, for the preparation of scientific reports.
2. The major sections of an APA paper include the title page, abstract, introduction, method, results, discussion, references, author note, tables, and figures.
3. The **title page** of a manuscript includes a manuscript page header and page number, the running head, the title, and author information.
4. The **abstract** provides a brief (100–120 words) summary of the contents of the paper.
5. The **introduction section** includes a thesis statement, literature review, and statement of the experimental hypothesis.
6. The **method section** contains a thorough description of the participants (**participants subsection**), the objects used in the experiment (**apparatus, materials, or testing instruments subsection**), and what took place during the study (**procedure subsection**).
7. The **results section** presents inferential and descriptive statistics to describe the experimental outcomes. **Figures** or **tables** may aid in presenting the statistical information.
8. In the **discussion section**, the researcher draws conclusions from the experiment by summarizing the results, comparing the results to previous research, and interpreting the results.
9. The **reference list** provides bibliographic information for any works cited in the paper. APA format includes different reference formats for periodical articles, books, chapters from edited books, and a host of other sources.
10. The **author note** allows the author to thank people for their help, cite a previous presentation of the findings, and designate a contact person for further information about the experiment.
11. The specific sections of the APA-format paper are designated by various headings. **Level 1** and **Level 3 headings** are most commonly used with experimental reports. For more complex papers, **Level 2** and **4 headings** may have to be added.

■ Check Your Progress

1. What is meant by the term *APA format*? Why was it developed?
2. Matching
 1. manuscript page header A. indicates the general topic you are interested in
 2. running head B. centered; upper- and lowercase letters
 3. level 1 heading C. first two or three words of the title
 4. level 3 heading D. left margin, underlined; upper- and lowercase letters
 5. level 4 heading E. condensed title
 6. thesis statement F. indented five spaces, underlined; only first word is capitalized; ends with period

3. Why is the abstract the most widely read section of most research reports?
4. How is the introduction section similar to what is found in a typical term paper? How is it different?
5. A citation includes three authors. The third citation of this study in the text should
 - a. list all three authors
 - b. list just the first author
 - c. list the first author followed by et al.
 - d. none of the above
6. List three different purposes of the method section. Which do you think is most important? Why?
7. We draw conclusions with _____ statistics and create a picture of our data with _____ statistics.
8. Could you use figures or tables as your sole information in a results section? Why or why not?
9. Some people believe that the discussion section is the most important section of an experiment report. Do you agree? Why or why not?
10. Why is the reference for a chapter from an edited book more complex than the reference for either a journal article or a book?
11. Matching:

1. title page	A. presents statistical findings
2. abstract	B. a short summary of the article
3. introduction	C. includes full address of author
4. method	D. reports the “bottom line” of the experiment
5. results	E. includes manuscript page header and running head
6. discussion	F. bibliographic information
7. references	G. reviews previous research
8. author note	H. tells how the experiment was conducted

Writing in APA Style

We hope that you were a good student in your English composition classes because good writing is quite important when it comes time to write your research report. Although we do not want to attempt to teach writing here, we do want to provide some helpful pointers. One of the topics covered in Chapter 2 of the *PM* is “expressing ideas.” You should read that whole chapter carefully. In the next sections we will give you some general and specific guidelines to assist you in writing your research report. As you will see, there are some differences between APA style and the way you learned to write in English classes. Scientific writing style is different in many ways from creative writing.

General Guidelines

The main objective of scientific writing is clear communication. It is your job to communicate your ideas as clearly as possible to the reader. The *PM* provides you with several hints about how to accomplish this goal; be sure to read pages 31–40 in the *PM* to supplement the following information.

Orderly Presentation of Ideas The key idea here is continuity. From the beginning of your research report to the end, you are writing a continuous idea or thought in which you tell the reader about your experiment. Do you remember sitting in class and getting lost in a lecture because you couldn't tell where the teacher was going (or where the teacher had been)? Those little side excursions in a lecture ("chasing rabbits") may be a delightful diversion in class, but they do tend to make you lose track of where you're going. Don't detour as you write your manuscript. Get on track and stay there with a single-minded purpose.

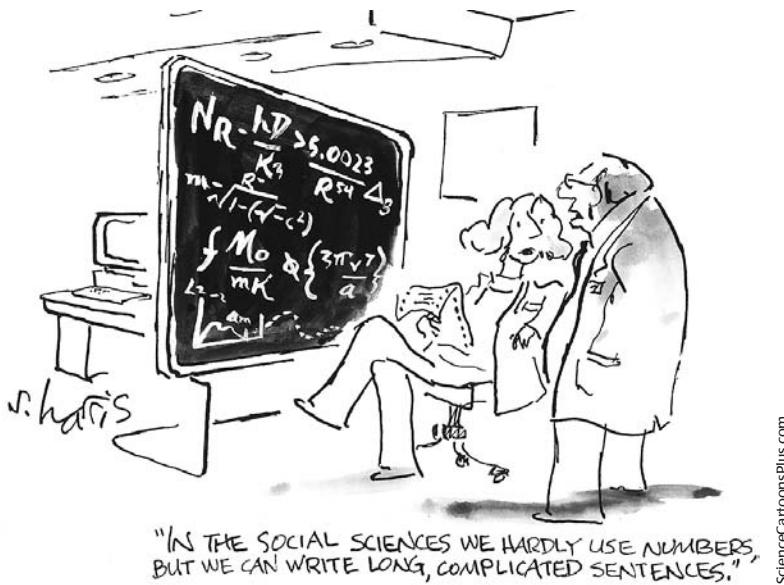
Smoothness of Expression Writing in a continuous fashion will greatly aid in your smoothness of expression. Creative writing often is not smooth because it uses literary devices to create tension or to conceal plot lines or to hide a surprise for later. Remember that scientific writing's goal is communication, not escape or entertainment. One of the best things you can do to make your writing smooth is to work on transition sentences when you shift from one topic to another. Try to avoid abrupt shifts that make readers feel they have run into a wall.

Economy of Expression Again, with your primary goal being communication, it is important to be direct and to the point in your writing. When journal editors work on submitted manuscripts, they have only a limited number of pages available for the printed journal; thus, it is to their advantage to have manuscripts that are short and communicative rather than long and unclear. Some people are surprised to find out that they can often shorten their writing and make it clearer at the same time. The *PM* specifically advises you to avoid jargon, wordiness, and redundancy. Also, you should not repeat yourself. (Yes, that second sentence was there on purpose—did you catch the unnecessary repetition?)

Precision and Clarity We encourage you to work on becoming a wordsmith rather than a wordmonger. As you probably know, a smith is someone who works with a particular material (e.g., a tinsmith, a goldsmith). A wordsmith works carefully with words, whereas a wordmonger uses words carelessly. Make sure the words you use fit the exact purpose and meaning that you have in mind. One of the major problems many of us have with writing is that we write in the same manner that we speak. Ambiguities may occur in speech, but we can clarify matters because we can interact directly with the speaker. Interaction is not possible when reading a text passage. Choose your words carefully so that you say just what you mean. Such clarity rarely occurs on a first attempt at writing: Reread and edit everything you write.

Strategies to Improve Writing Style The *PM* (APA, 2001, p. 40) suggests three approaches to becoming a more effective writer.

1. **Write from an outline.** If you have a "road map" for your writing, you are more likely to arrive at your destination in a timely fashion.
2. **Write your first draft, put it away, and reread it after a delay.** If you attempt to cram your writing into a short time period, you will have difficulty editing your writing because you are likely to have the same thoughts you had a few minutes earlier. By giving yourself a time break, you are more likely to see the things you missed the first time—and it will be easier to think of ways to correct the problems.



sciencecartoonsplus.com

We hope you recognize that both statements in this cartoon are incorrect. In psychology, we do use numbers and we work hard at communicating clearly.

3. **Ask someone to evaluate your writing.** It is usually quite helpful to have at least one other person read your papers before you submit them. It is easier for someone who is unfamiliar with the work to spot inconsistencies, weaknesses, ambiguities, and other flaws in your writing. Of course, you are asking *only* for a critique of your writing style, not the substance of your paper. Having someone help you with substance is unethical unless that person is a coauthor. Some instructors offer to critique rough drafts of your work—you should always take advantage of such an offer. Ask classmates to critique your writing and offer to do the same for them. You may learn a great deal both from their critiques and from reading and critiquing someone else’s writing.

Grammatical Guidelines

The *PM* covers a variety of guidelines about grammar on pages 40–61. Most of these guidelines are standard conventions of grammar that you learned in grammar classes. We urge you to review these pages. Rather than turning this book into a grammar handbook at this point, we will cover only those conventions that are specific to APA style or with which students seem to have difficulty.

Passive Voice According to the *PM*, as well as other style guides, you should use active voice rather than passive voice in writing your research report. In passive voice, the true object of the verb becomes the subject of the sentence and the true subject becomes the object (Bellquist, 1993). Passive voice often appears in methods sections because writers use the passivity to avoid personalizing that section. Let us give an example to clarify:

After viewing the slides, a recall test was given to participants.

This sentence is not direct and active; rather, it is indirect and passive. The test should be the object of the sentence, not the subject. Who did the acting? Presumably the experimenter did, but the experimenter is not even present in this sentence.



Reread the passive voice sentence. Can you recast it in active voice?

Actually, there are several ways to change this sentence into an active voice, depending on whether you want to include the experimenter. You could write,

I gave the participants a recall test after they viewed the slides.

If you have coauthors for your experiment, the sentence could become this:

We gave the participants a recall test after they viewed the slides.

Although many experimenters seem to be uncomfortable using first person (*I, we*), the PM does permit it. In fact, the manual specifically says to avoid third-person references to yourself (e.g., “the experimenter,” pp. 37–38). If you still wish to avoid first person, you could write,

The participants completed a recall test after viewing the slides.

In each of these sample sentences, you now have actors acting (active voice) rather than having something done to them (passive voice).

That Versus Which Clauses beginning with *that* are termed restrictive clauses and should be essential to the meaning of the sentence. Clauses beginning with *which* can be either restrictive or nonrestrictive (simply adding additional information). In APA style, you should confine yourself to using *which* for nonrestrictive clauses. Thus, you should not use *that* and *which* interchangeably. Using *which* is similar to making an “oh, by the way” addition to your sentence. To further help you distinguish the difference, remember that nonrestrictive clauses should be set off with commas. Let’s look at an example:

The stimulus items *that* the participants did not recall were the more difficult items.

The phrase “*that* the participants did not recall” is essential to the sentence’s meaning. Imagine the sentence without that phrase—it would make no sense. Let’s look at another sentence:

The stimulus items, *which* were nouns, appeared on a computer monitor.

The phrase “*which* were nouns” is not essential to the meaning of this sentence. If we delete this phrase, the sentence retains its original meaning. The phrase does add some additional information about the stimulus items, but you could include this information elsewhere.

Words With Temporal Meaning The words *since* and *while* can cause difficulty in scientific writing because they have more than one meaning in everyday usage. Writers often use *since* interchangeably with *because* and use *while* to substitute for *although*. Some grammar

books allow these multiple uses. APA style, however, does not. You should use *since* and *while* only for temporal purposes—in other words, to make time comparisons. Thus, use *while* to denote events that occur at the same time and *since* to denote that time has passed. Again, here are some examples:

Many different IQ tests have evolved *since* Binet's original version.

Note that here *since* refers to time that occurred after Binet's test; this sentence is acceptable.

Since the XYZ group scored higher, we concluded that they learned the material better.

This use of *since* is incorrect; you should substitute *because* in its place.

While the participants were studying the verbal items, they heard music.

Note that *while* in this sentence tells you that studying and music playing occurred simultaneously; this sentence is acceptable.

While some psychologists believe in Skinner's ideas, many others have rejected his beliefs.

This use of *while* is incorrect—nothing is occurring at the same time. Instead, a contrast is being drawn. You should substitute *although* or *whereas* in this sentence.

Bias in Language Remember that we stressed the use of unbiased language earlier in this chapter. We believe that this type of writing is important in helping maintain a neutral (unbiased) approach to science. Thus, we wish to remind you of the need for removing biased terms from your writing. The *PM* gives three guidelines that may be helpful in reducing bias in writing (APA, 2001, pp. 61–76).

- **Describe at the appropriate level of specificity.** In other words, you should describe people as specifically as you can. When we use broad terms to describe people, we are more likely to include people who should not be included. For example, “Japanese Americans” is more specific than “Asian Americans.”
- **Be sensitive to labels.** When we use stereotyped labels, we are likely using terms that contain bias. Basically, we should refer to groups as they wish to be referred to rather than imposing our labels on them. When at all possible, it is better to avoid labels. As the *PM* points out, “people diagnosed with schizophrenia” is both more accurate and more preferred than “schizophrenics” (APA, 2001, p. 64).
- **Acknowledge participation.** This guideline is generally aimed at experiments that use human participants, although it would not hurt us to keep it in mind for animal studies also. The general idea of this guideline is to make sure you remember that the participants in your experiment are individuals. This idea formed the rationale for changing the label *subjects* to the label *participants*. Using active rather than passive voice also helps to personalize your participants.

A Disclaimer Please remember that we could not possibly squeeze all the grammar guidelines from the *PM* into this section. Again, we chose to highlight the few that we did because they may differ from what you learned in English classes or because we know that students (and professors) tend to have trouble with these points of grammar and usage. We did not leave out the others because they are unimportant or even less important. We urge you to read pages 40–61 in the *PM* to review your knowledge of grammar.

APA Editorial Style

Chapter 3 of the *PM* addresses APA editorial style on pages 77–214—virtually a third of the book. This chapter gives writers a style guide to follow that will help create uniformity in writing by different authors in different publications. We have already covered the most important aspects of APA editorial style in this chapter: levels of headings, metrification, statistical copy in text, tables, figures, reference citations in text, and reference lists.

In addition to these important aspects of APA editorial style, you should be aware that the *PM* gives you guidance on such issues as punctuation, spelling, capitalization, italics, abbreviations, seriation, quotations, numbers, footnotes, and appendixes. Again, we do not have the space it would take to address every possible concern in this chapter. When you have questions about any of these matters, consult Chapter 3 in the *PM*.

Preparing Your Manuscript

Chapter 5 of the *PM* (pp. 283–320) provides the guidelines you need in order to type your experimental paper. This is probably one of the most-used chapters in the manual because it includes three sample papers (pp. 306–320). These sample papers include notations of specific *PM* sections for each important component of the paper. We hope that the combination of the sample manuscript in this chapter and the sample papers given in the *PM* will make typing your paper a relatively simple matter.

Chapter 5 of the *PM* is primarily a reference chapter much like Chapters 3 and 4. You should consult it whenever you have a question about typing a specific portion of your manuscript. Let us provide you with a short list of the highlights of the typing instructions:

- **Line spacing.** Double-space everything everywhere.
- **Margins.** Use at least 1-inch margins on all sides. Keep in mind that for the top margin, this figure refers to the point at which the text begins rather than the manuscript page header and number. Thus, you can set the top margin in your word processor to less than 1 inch so that the text begins at least 1 inch down from the top of the paper.
- **Lines.** Set your word processor to left justification. Your paper should have a ragged right edge throughout (i.e., the right margin should *not* line up down the page like it does in this book). Do not divide or hyphenate words between lines.
- **Pages.** Number all pages (including the title page; excepting figures) consecutively. The following sections should begin on new pages: title page, abstract, introduction (remember not to label it *Introduction*), references, appendixes, author note, footnotes, tables (a separate page for each), figure captions (all on one sheet), figures (each on a separate page).
- **Word spacing.** Space once after all punctuation, including plus and minus signs in equations. There is no spacing before or after hyphens (-) or dashes (– or —). Use a hyphen to denote a negative value; in this case use a space before the hyphen but not after.
- **Quotations.** Enclose quotations that are shorter than 40 words in double quotation marks (" ") and write them as part of the text. You should block (indent) longer quotations from the left margin—be sure to double-space them.

Consult Table 14-2 for more comprehensive guidelines regarding APA formatting of your manuscript. If you have questions about other matters as you type your manuscript, consult Chapter 5 of the *PM*.

TABLE 14-2 Checklist for Correct APA Formatting**General Formatting and Typing***Publication Manual Section*

- There are 1-in. (2.54-cm) margins on all four sides of each page of the manuscript. 5.04
- The typeface is the correct size (12 points on a word processor) and the correct style (serif typefaces such as Courier or Times Roman). 5.02
- The manuscript is double-spaced throughout, including title page, references, tables, figure captions, author notes, and appendixes. 5.03
- The page header is the first two or three words of the title. 5.06
- The page number appears (a) on the same line with the page header and is five spaces to the right or (b) below the page header. 5.06
- The page header and page number are typed at the top of each page of the manuscript (except pages containing figures). 5.06
- The page header and figure number are handwritten on the back of figures (or on the front, outside the image area of the figure). 5.22
- There is only one space after these punctuation marks: commas, colons, semicolons, punctuation at the end of sentences, periods in citations, and all periods in the references section. 5.11
- URLs should work; in Reference list they should have no end punctuation. 4.15, 4.16
- Lowercase letters in parentheses have been used to indicate a series of events or items within a paragraph. 5.12
- Words are not broken (hyphenated) at the end of a line. 5.04
- All units of measurements have correct abbreviations. 3.25, 3.51
- Arabic numbers have been used correctly to express all numbers in the abstract 3.42
numbers that are 10 or larger
numbers less than 10 *only when* those numbers are compared to a number greater than 10 (e.g., "Participants included 15 humanities and 3 natural science majors.")
numbers that immediately precede a unit of measurement
numbers that represent fractions and percentages
numbers that represent times, dates, ages, participants, samples, populations, scores, or points on a scale
- Words have been correctly used to express 3.43
numbers less than 10
numbers at the beginning of a title, sentence, or heading

(Continued)

TABLE 14-2 (Continued)

Publication Manual Section

Title Page

The running head is aligned with the left margin and is equal to or less than 50 characters and spaces long.

1.06, 5.15

The author note *does not* appear on the title page; instead, the author note appears on a separate page after tables, figures, and appendixes (if included).

3.89, 5.20

Abstract

The word Abstract is typed at the top of the page.

5.16

The first line of the abstract is even with the left margin.

5.16

The abstract is not more than 120 words.

1.07, 5.16

Body of the Manuscript

There are *no* one-sentence paragraphs.

2.03

Gender-inclusive language is used through plural pronouns (e.g., *they, their*), by using nouns (e.g., *one, an individual, participant's*), by sparse use of *he or she* or *she or he*, or by sparse use of alternating between gendered pronouns (e.g., *he . . . ; she . . .*).

2.13

The words *male* and *female* are used *only* as adjectives (e.g., *female quail*), whereas the words *men, women, boys, and girls* are used as nouns.

Table 2.1

Quotations are word-for-word accurate and page numbers are provided.

3.35, 3,39

The word *while* is used *only* to indicate events that take place simultaneously (alternatives: *although, whereas, and, but*).

2.10

The word *since* is used *only* to indicate the passage of time (alternative: *because*).

2.10

Terms that are abbreviated are written out completely the first time they are used, then always abbreviated thereafter.

3.21

Latin abbreviations are used sparingly and *only* in parenthetical material.

3.24

The word *and* is used in citations outside of parentheses.

3.95

The ampersand (&) is used in citations within parentheses.

3.07, 3.95

When two or more citations are in parentheses, the citations are typed in the same order in which they appear in the references section.

3.99

(Continued)

TABLE 14-2 (Continued)

<i>Publication Manual Section</i>	
Each and every citation used in the manuscript is correctly typed in the references section.	4.01, 4.02
The phrase <i>et al.</i> is used with each citation that lists six or more authors, and with each citation that lists three to five authors <i>after</i> the first instance of that citation.	3.95
In the method section the word <i>participants</i> is consistently used (use <i>subjects</i> only with animals).	1.09, 2.12
In the results section all test statistics (e.g., <i>F</i> , <i>t</i> , <i>p</i>) are italicized.	3.19, 3.58
References Section	
All entries are typed in alphabetical order.	4.04
Each and every entry occurs in the body of the manuscript.	4.01
Authors' names are separated by commas.	4.08
The volume numbers of journals are italicized.	4.11
Each entry is typed in a "hanging indent" format, meaning that the first line of each reference is flush with the left margin and subsequent lines are indented.	4.07, 5.18
The names of journals, book chapters, and books are correctly capitalized.	4.11

Source: Adapted from Dunn, J., Ford, K., Rewey, K. L., Juve, A., Weiser, J. A., and Davis, S. F. (2001). "A Modified Presubmission Checklist." *Psi Chi Journal of Undergraduate Research*, 6, 142–144. Copyright © 2001 Psi Chi, The National Honor Society in Psychology (www.psichi.org). Reprinted by permission. All rights reserved.

Student Views of Professional Activities

Preparing an APA-format paper does indeed involve a reasonable amount of effort; however, the rewards for this effort are genuine and substantial. You will have mastered a valuable skill, writing in an accepted scientific format, and you will have a completed paper that you can submit for publication. Throughout this book we have "gently encouraged" you to take advantage of professional development opportunities such as presenting your research at a convention and publishing your research. Let's conclude with a more in-depth look at these professional activities. They are not as daunting as you might think.

Your first decision will be which type of professional activity to pursue. In Chapter 1 we talked about sharing your results through presentation and publication. In this chapter we have primarily focused on publication with Dickson's paper. You may, of course, have two different types of presentation opportunities: presenting your research in either a poster or paper format at a conference (see Chapter 1 for a list of possible conferences). In a poster format, you

will display your research in a visual format—most often a small amount of text to inform viewers about the basics of your experiment plus graphs, figures, or charts to depict your findings. Typically, poster sessions take place in a large room with quite a few researchers displaying posters simultaneously. Conference attendees walk through the room to view the posters, often stopping to converse with researchers whose posters interest them. They may have questions for you or offer suggestions for further research on your topic.

On the other hand, a paper presentation consists of making an oral presentation of your research. You might have 10–15 minutes to talk about your research and findings in a room where people are seated. Typically, you would give a brief summary of the background literature, your methodology, your findings, and the implications of your research—much like the order of the sections in your written paper. Researchers presenting talks often display their findings visually with overhead transparencies or computer projection systems. Usually, the audience has a few minutes after your presentation to ask questions.

If you plan on attending a conference to make either a poster or paper presentation, you typically will have to submit an abstract of your presentation to the conference organizers for their review. The organizers will choose to accept or reject each abstract submitted. Depending on the type of conference, the rate of acceptance will vary; typically, conferences designed specifically for undergraduates will have high rates of acceptance. Professional conferences designed more for graduate students and faculty often have lower rates of acceptance. Because of the review process, conference presentations are considered professional activities; you can record such activities on your vita (an academic version of a résumé, see Landrum & Davis, 2007). Researchers often make a presentation of their research at a conference before attempting to publish that research. The feedback you receive from an audience might help you improve your manuscript before you pursue publication.

If you decide to try to publish your research, your faculty advisor can assist you in deciding where to submit your paper (we made some suggestions in Chapter 1). Once you have made that decision, you should consult a recent issue of the journal you have selected to determine exactly how many copies to send, as well as any other special requirements the journal may have. Follow these instructions very carefully. When your submission is received, the editor, most likely, will send your paper to several (usually three) reviewers who will read, critique, and evaluate your work. After the reviewers have finished reading and critiquing your paper, each one will make a publication recommendation (such as accept, accept pending minor revisions, accept pending major revisions, or reject) to the editor. The editor will evaluate the reviewers' recommendations and comments and make a final publication decision about your paper. The editor's letter describing the publication decision, as well as the reviewers' comments and critiques (often including the copies of your paper that they have marked and returned to the editor), is then returned to you. This process can take two or more months, so be patient.

It is likely that you will have to revise your paper one or more times before the editor and reviewers determine it is suitable for publication. Don't give up! If you are interested in understanding the publication process from the editor's point of view, we recommend that you read Smith's (1998) article. Consult with your faculty advisor at each step of the process. Having a publication is well worth the time and effort you will expend.

Very briefly, those are our views on these professional activities; we have asked three students to give you their views about them. We turn now to the students' views of making a poster presentation, presenting a paper, and publishing an article.

Presenting a Poster at an Undergraduate Psychology Conference

I spent most of my undergraduate college career as a nontraditional student, meaning older than the other students and working a full-time job. When I began to reach the end of my undergraduate studies and really started thinking about graduate school, I found out quickly that with the ratio of applications to acceptances given for my chosen field, I would need something to set myself apart from the other applicants.

I began taking advantage of the advising sessions offered at my university and found out that I could do research projects with my professors. These research projects included gaining IRB approval, setting up surveys, informed consents, debriefing forms, administering surveys, setting up the database, and data analysis. I found immediately that I loved research!

One project I developed was entitled *Sexual Orientation and Family Dynamics: Exploration of the Psychosocial View*; it was based on my interest in prejudicial attitudes. I developed this project while I was in a Human Sexuality class. There was a passage in my textbook that mentioned the psychosocial view of sexuality, which is basically the idea that nurture causes sexuality to develop. Two panels of Lesbian, Gay, Bisexual, and Transgender (LGBT) individuals spoke to our class; they discussed their development and talked about their feelings and the reactions of family and friends. I noticed a common theme with almost every speaker. They all stated that they knew very early in childhood (between the ages of 3 and 5) that they were “different” from their peers. This information made me question the psychosocial theory.

Once I had developed a working survey, I completed the paperwork for IRB approval. This process went smoothly. Most of the professors in the psychology department at my university encouraged their students to participate, and I soon had over 200 participants from my department alone.

I was extremely proud of this research; when the results showed something new, I wanted to share it. My faculty mentors encouraged me to present my research at a regional conference for undergraduates in my area. I submitted my application and received approval to present a poster within a week. My next challenge was finding what information needed to be displayed in this session. My mentors suggested that I limit my poster to 10–12 PowerPoint sheets for the sake of clarity. I found that the hardest part of the entire process was picking and choosing the most important parts. Finally, I cut my presentation down from 37 to 12 PowerPoint sheets.

My poster session was presented at the Georgia Undergraduate Research in Psychology (GURP) Conference; this conference is held for all universities in the Southeastern U.S. One note that I would like to add here is that many universities have monies set aside for students to attend and present oral papers and posters at conferences. I encourage you to explore the possibility of obtaining university funding; attending a conference can be expensive. I was fortunate enough to have professors who gave me this information and helped me obtain the funding I needed.

The morning of the conference, I think I changed clothes about 10 times. I have given presentations in class—my position as a teacher’s aide allowed me to give lectures; however, poster sessions are different. Your material is set up on a poster board. The people at the conference are other psychology students and professors. These

attendees walk around and read the posters that interest them. Hence, you give mini-lectures about your poster to ever-changing groups. It still seems odd to me that I was “teaching” my professors about a topic that they had not previously studied.

The experience of going to GURP was very rewarding. I was able to expand my knowledge in an area that interested me. I was also able to share my findings with others who were interested in the same topic area. This conference allowed me to network with people who are currently at universities that I was researching as potential graduate schools. These conversations helped me to narrow my focus and refine my list of potential universities, as well as giving me very important tips for my applications. This experience gave me a lot of important knowledge for my continued education, allowed me to network, get comfortable in front of crowds, and guide the direction of my future pursuits.

Deborah J. Kemp
Kennesaw State University

The Catalyst for Becoming a Professional Psychologist

When pursuing a career in psychology you will face many challenging tasks. Undergraduates constantly learn new information, take tests, and strive to succeed. One challenge many undergraduates face is giving an oral presentation for the first time. This task can seem overwhelming and intimidating. However, with a few steps and guidelines to help along the way, I believe any willing student can succeed in this task. In the end, all the hard work will pay off when the once unbearable oral presentation turns into a catalyst for a successful undergraduate career in psychology.

I was first given the intense but bearable task of giving an oral presentation my sophomore year in college. I enrolled in an honors section of a research and statistics class and was given the assignment of conducting a literature review to support my research project. This literature review eventually became a presentation to be given orally at an undergraduate psychology conference. Because I had never conducted any research, performed a literature review, or given an oral presentation, this task was most challenging. I picked my oral presentation topic to support my research project on the effect of clothes and misinformation on test scores. The final topic I agreed to present was experimenter effects. I found this topic of specific interest because of its constant threat to all experimental research. I began my literature review by accumulating articles to support the fact that experimenter effects exist; I found myself swimming in a sea of information. I did not know how to condense all this useful information into one cohesive oral presentation. Moreover, my oral presentation was limited to 12 minutes. During this time I had to summarize a vast amount of data and convince the audience that experimenter effects exist and are very important. The task of pulling together all my information and summarizing the data into a short set of PowerPoint slides set the stage for my oral presentation at the 2005 Great Plains Students’ Psychology Convention held in Omaha, NE.

I believe the most challenging part of making an oral presentation is speaking to the audience. This comment may sound like simple stage fright, but in reality the hard part is speaking to your audience and not reading your notes or PowerPoint slides. In order to present successfully, the speaker must know the information so that

he/she can explain the topic instead of simply reading to the audience. Some advice from my faculty mentor helped me learn how to explain and not read my paper. She told me to think of presenting orally as teaching, not presenting information. When thinking of an oral presentation as a lecture I was able to have more confidence, because I thought of myself as a teacher. When teaching a subject the lecturer thinks that he/she knows more about the subject matter than the audience. This added confidence helps the presenter speak to the audience instead of reading to them. Presenting/lecturing to my audience took some practice. First I practiced in front of my faculty mentor. By this time I had already read my information numerous times in order to make my PowerPoint slides. All I needed to do was get in front of my professor and talk to her as though I was teaching her about what I had read and learned. I was amazed at how confident I was at knowing my information. The only problem I had was the 12-minute time limit; I went over time. So I had to go home and practice more, so that I could get my time down, while still covering all the relevant information. This task was very similar to my earlier task of condensing information from my literature review to fit into a PowerPoint. I then practiced in front of my peers, and they helped me to see portions of presentation that were not as understandable as the rest of the talk. Then I practiced about five times in front of my peers and faculty mentor and numerous times to myself. With each practice I gained a bit more confidence.

Going to the Great Plains Conference was an intensely enjoyable experience. I spent time with my friends, met new influential individuals, and gained knowledge about my favorite field: psychology. When I first arrived at the conference I needed to prepare for my presentation. I found the room I was to present in, downloaded my PowerPoint slides, and waited for my time to speak. When I first stood in front of my professors, peers, influential professionals, and session judges, I was nervous. I shook at the thought that I could mess up in front of all these people who had a direct bearing on my future in psychology. However, as I began my lecture and did make a mistake, I took a moment to laugh at myself, and that helped me to calm down. I realized that the people in the audience were listening, not judging. The audience is there because they support you; they hope you will succeed. I began to relax, and I taught my audience all I knew about experimenter effects. My first oral presentation was over; I had survived!

I was thrilled when it was announced that I had won the first-place award in my session. In addition to this recognition I made new professional relationships and grew professionally by attending the conference. I believe that you can have just as much success making an oral presentation if you stay organized, practice, and have confidence in yourself. Making an oral presentation can act as a catalyst for a successful career in psychology.

Christine Yates
Emporia State University

In Chapters 4 and 8 you read about Carolyn Licht's research on occupational stress. Here are Carolyn's views on the publication process.

The Dance of the Researcher

The publication process is both a frustrating and exhilarating experience. It takes commitment, perseverance, and discipline to make it to the culminating moment in

which you see your work in print. When that moment happens, all the time and effort expended is forgotten and all you feel like doing is getting up and dancing a jig.

I had my first opportunity to discover the exhilaration associated with conceptualizing, analyzing, and preparing my research for presentation and publication as a sophomore in college. I had never done any formal research of any kind before, having been a ballet dancer for the previous decade of my life, and was a bit overwhelmed when my first assignment for my Experimental Psychology was to create and execute a project from scratch. I decided to pick a subject that was something I and others could easily relate to—occupational stress. If I had known how consuming the research would become over the next few years, I probably would have picked a less “stressful” topic. This school assignment opened my eyes to the limitations, obstacles, and rewards faced by researchers trying to develop a workable hypothesis, gather and analyze data, and present results.

What I found to be the most interesting aspect about the research and publication process, ironically, was how closely it paralleled the work I did in my previous career as a professional ballet dancer. The ultimate goal and satisfaction of both endeavors is the sharing of your labor with others outside of your immediate circle. The journey that must be taken to get to this point is full of frustrations, obstacles, criticisms, and excitement; the cliché “No pain, No gain” seems apropos to characterize the overall process. The “dance” begins with an idea that takes form as the dancer/researcher collects the data (choreography and music/literature review and survey results), begins the rehearsal process, and analyzes all the components to see how they come together. The frustration of realizing that the pieces are not flowing together as smoothly as predicted, such as when you send out 500 surveys and receive only 100 completed ones back, is reminiscent of discovering that you have run out of choreography before the music is over. Ultimately you must return to the drawing board and either collect more data or reconceptualize the original premise of the piece. Eventually, however, a moment arrives in which what once was only an idea in your mind becomes a reality, although the form it takes may be far from what you had originally conceived.

The pride you feel when you have completed the research and submitted it to a journal for publication is similar to the joy you feel when you perform your dance for the first time on stage. You learn quickly, however, as you receive your first critiques from the journal reviewers, that you are still in dress rehearsal and are far from being ready for the final performance. I remember vividly the crushing feeling I got when I opened the large manila envelope containing the four drafts of my “perfect” paper, only to discover the massive number of red marks and comments concerning everything that was wrong and needed to be addressed before my work would even be considered for publication. The fact that many of the various suggestions from the reviewers were contradictory (one loved a section of the paper while another said it should be deleted altogether) only added to my feelings of despair. Gratefully in the midst of all the criticism was also the glimmer of hope, the encouragement I needed to “rehearse” a bit longer, which consisted of a simple checked box on the comment sheets from the reviewers; the box marked was “accepted pending revisions” as opposed to the one next to it that said “rejection.” With less work than I had anticipated, the next box to be checked would be the one marked “accepted for publication”; all the effort and energy

exerted no longer seemed at all troublesome as the excitement of the pending publication became the focus of my attention.

When I saw my name in print, the feeling of pride and the boost to my self-esteem was incredible, but as I read my article in the journal, I realized that this was just the first step in what was going to be a lifelong journey in the field of psychology. I looked at my paper and understood that although my results were fascinating, they were limited because of the nature of the question and the research design. I wanted—no, I needed to discover if my findings were true on a larger scope, with a diverse population, in different settings. I knew at that moment what so many researchers before me have learned, that research questions are never definitively answered; rather, they simply raise more questions that stimulate the researcher to continue the journey exploring the path ahead.

Ultimately, this college experimental research assignment resulted in two publications in the *Psi Chi Journal of Undergraduate Research* (Licht, 2000; Licht & Solomon, 2001), a poster presentation at the Eastern Psychological Association's National Convention in Baltimore, an oral presentation at the Honors Colloquium at my college, citations in two textbooks, and, most importantly, it was a substantial contributing factor to my decision to pursue and eventual acceptance into a doctoral program in Clinical Psychology. In addition, I discovered, through my numerous interactions with the editor of the journal, that the publication process is a way to connect with others in the field, and to develop professional relationships and true friendships that become the true rewards of taking this particular journey.

Carolyn Licht
Marymount Manhattan College

We hope that you, too, will have the opportunity to have the experience of presenting your research as Deborah and Christine did and publishing your manuscript like Carolyn. Although there is hard work involved, the payoff is worth it all. We will remind you of the publication opportunities we mentioned in Chapter 1. As we have stressed throughout this book, publishing is not something that is beyond your capabilities. If you never give it a try, you certainly can't succeed. Good luck!

■ REVIEW SUMMARY

1. The primary goal of scientific writing is clear communication.
2. Goals that aid in clear communication are orderly presentation of ideas, smoothness of expression, economy of expression, and a striving for precision and clarity.
3. To improve your writing style, you should write from an outline, put away your first draft before editing it, and have someone evaluate your writing.
4. You should use active voice whenever possible in writing your research report.
5. *That* should be used only with restrictive clauses, which include information that is essential to the meaning of a sentence. *Which* should be used in nonrestrictive clauses, which add information but are not essential to a sentence's meaning.

6. *Since* should not be used to substitute for *because*, nor should *while* substitute for *although*. Both *since* and *while* should be used only for temporal (time-related) meaning.
7. Psychologists strive to use unbiased language in their writing.
8. APA style includes guidelines on such diverse matters as punctuation, capitalization, quotations, numbers, appendixes, and typing guidelines.
9. The *Publication Manual of the American Psychological Association* (2001) is the stylebook for psychological writing. It contains a wealth of information about the writing process.

Check Your Progress

1. What would be wrong with writing your research paper in the style of Twain, Hemingway, or Faulkner? Be as specific as possible in your answer.
2. What are the three strategies to improve your writing style? As you list each strategy, also tell what you would have to change about your writing style to incorporate the strategy.
3. Which of the following illustrates passive voice?
 - a. The experimenter gave the memory test to the participants.
 - b. The participants took the personality test after a rest period.
 - c. The endurance test was given by the experimenter's assistant.
 - d. All of the above.
 - e. None of the above.
4. Change each of the following sentences in passive voice to active voice.
 - a. An experiment was conducted by Jones (1995).
 - b. The participants were seated in desks around the room.
 - c. The stimulus items were projected from the rear of the cubicle.
 - d. A significant interaction was found.
5. Choose the correct sentence from each pair below. Add punctuation if necessary. Explain your answers.
 - a. The experimenter tested the animals which were older first.
The experimenter tested the animals that were older first.
 - b. A room which was a classroom was the testing environment.
A room that was a classroom was the testing environment.
6. Decide whether each sentence below is correct or incorrect. If it is incorrect, correct it.
 - a. Since you are older, you should go first.
 - b. Since I began that class, I have learned much about statistics.
 - c. While we are watching TV, we can also study.
 - d. While you are older than I, I should still go first.

7. Use unbiased language to express each phrase.

Orientals	elderly
mankind	girls and men
mothering	chairman
homosexuals	depressives

8. Correct the following incorrect expressions.

$a + b = c$	trial - by - trial
-1	Enter: Your name

■ Key Terms

APA format, 333	Reference, 342	Procedure
Headings, 334	Reference section, 342	subsection, 346
Title page, 335	Unbiased language, 343	Results section, 348
Manuscript page header, 335	Method section, 343	Figure, 350
Running head, 335	Participants subsection, 343	Table, 350
Level 1 heading, 336	Level 3 heading, 343	Discussion section, 353
Abstract, 337	Materials subsection, 344	Plagiarism, 356
Introduction, 338	Apparatus subsection, 346	Author note, 362
Thesis statement, 338	Testing instrument(s)	Level 4 heading, 363
Citation, 342	subsection, 346	

■ Looking Ahead

At this point we have reached the end of this text—there is no Chapter 15. We do, however, look ahead to your research career. Perhaps your research career will be nonexistent; you may not be required to design, plan, and conduct an experiment as part of this course or another course. In this case we hope you have learned something about research that will make you a critical consumer of research information in the future. Perhaps your research career will entail only one study—the one you conduct for this course. We believe this book will prove helpful for you in that endeavor. Finally, perhaps some of you now envision an ongoing research career for yourselves. We hope this book has opened your eyes to the powerful possibilities of experimental research in psychology and that you are eager to follow that path in the future.

Regardless of what your future plans regarding research are, we hope we have made you think, challenged you to work, helped you contemplate conducting research, and perhaps entertained and amused you a little along the way. All of you will be faced with research in some fashion in your future. We wish you luck as you begin your journey.

Glossary

A Refers to the baseline measurement in a single-case design.

A-B design A single-case design in which you measure the baseline behavior, institute a treatment, and use a posttest.

A-B-A design A single-case design consisting of a baseline measurement, a treatment, a posttest, and a return to the baseline condition. It may not be recommended if the participant is left without a beneficial or necessary treatment in the second baseline.

A-B-A-B design A single-case design consisting of a baseline, treatment, posttest, return to baseline, repeated treatment, and second posttest. This design gives the best chance of isolating causation.

Abscissa The horizontal or x axis of a graph.

Abstract A brief description of the research that is presented in an APA-format paper.

Achievement test Designed to evaluate an individual's level of mastery or competence.

Analytic statement Statements that are always true.

Analytic survey Seeks to determine the relevant variables and how they are related.

APA format Accepted American Psychological Association form for preparing reports describing psychological research.

Apparatus subsection The second subsection of the method section. When appropriate, contains information about the equipment used in an experiment.

Aptitude test Designed to assess an individual's potential ability or skill in a particular job.

Asymptotic Refers to tails of distributions that approach the baseline but never touch the baseline.

Author note A note at the end of an experimental report that contains information about the author or paper for readers' use.

Axial coding The process of rearranging data after open coding so that new relations are formed between concepts.

B Refers to the outcome (treatment) measurement in a single-case design.

Balancing A control procedure that achieves group equality by distributing extraneous variables equally to all groups.

Bar graph A graph in which the frequency for each category of a qualitative variable is represented as a vertical column. The columns of a bar graph do not touch.

Baseline A measurement of a behavior made under normal conditions (i.e., no IV is present); a control condition.

Between-groups variability Variability in DV scores that is due to the effects of the IV.

Between-subjects comparison Refers to a contrast between groups of participants who were randomly assigned to groups.

Carryover effect The effects of one treatment persist or carry over and influence responses to the next treatment.

Case studies Studies involving the intensive observation of a single participant over an extended period of time.

Case-study approach An observational technique in which a record of observations about a single participant is compiled.

Cause-and-effect relation Occurs when we know that a particular IV (cause) leads to specific changes in a DV (effect).

Citation A notation in text that a particular reference was used. The citation provides the

name(s) of the author(s) and date of the work's publication.

Cohort A group of individuals born during the same time period.

Comparative psychology The study of behavior in different species, including humans.

Complete counterbalancing All possible treatment sequences are presented.

Completely randomized ANOVA This one-way ANOVA uses independent groups of participants.

Concurrent validity Degree to which the score on a test or inventory corresponds with another measure of the designated trait.

Conditional matrix A diagram that helps the researcher consider the conditions and consequences related to the phenomenon under study.

Confirmability The extent to which the qualitative research report is accurate, unbiased, and can be confirmed by others.

Confounded experiment An experiment in which an extraneous variable varies systematically with the IV, which makes drawing a cause-and-effect relation impossible.

Confounding A situation in which the results of an experiment can be attributed to either the operation of an IV or an extraneous variable.

Constancy A control technique by which an extraneous variable is reduced to a single value that is experienced by all participants.

Content validity The extent to which test items actually represent the type of material they are supposed to represent.

Contradictory statement Statements that are always false.

Control Either directly manipulate (1) a factor of interest in a research study to determine its effects or (2) other, unwanted variables that could influence the results of a research project.

Control group In a two-group design, the group of participants that does not receive the IV.

Control procedure One of several steps experimenters take to ensure that potential extraneous variables are controlled, including random assignment, matching, and so on.

Convenience sampling A researcher's sampling of participants based on ease of locating the participants; often it does not involve true random selection.

Correlated assignment A method of assigning research participants to groups so that there is a relationship between small numbers of participants; these small groups are then randomly assigned to treatment conditions (also known as *paired* or *matched assignment*).

Correlated groups Groups of research participants that are related in some way, for example, matching, repeated measures, or natural sets.

Correlation coefficient A single number representing the degree of relation between two variables.

Correlational study Determination of the relation between two variables.

Counterbalancing A procedure for controlling order effects by presenting different treatment sequences.

Credibility The accuracy of the identification and description of the subject of the study.

Criterion validity Established by comparing the score on a test or inventory with a future score on another test or inventory.

Cross-cultural psychology A branch of psychology whose goal is to determine the universality of research results.

Cross-sectional research Comparison of two or more groups during the same, rather limited, time period.

Cultural response set The tendency of a particular culture to respond in a certain manner.

Culture Lasting values, attitudes, and behaviors that are shared by a group and transmitted to subsequent generations.

Debriefing session The time at the conclusion of an experiment when its nature and purpose are explained to participants.

Deductive logic Reasoning that proceeds from general theories to specific cases.

Degrees of freedom The ability of a number in a specified set to assume any value.

Demand characteristics Features of the experiment that inadvertently lead participants to respond in a particular manner.

Demographic data Information about participants' characteristics such as age, sex, income, and academic major.

Dependability The extent to which the researcher believes that the same results would be produced if the study were replicated.

Dependent variable (DV) A response or behavior that the experimenter measures. Changes in the DV should be caused by manipulation of the independent variable (IV).

Descriptive research methods Research methods that do not involve the manipulation of an independent variable.

Descriptive statistics Procedures used to summarize a set of data.

Descriptive survey Seeks to determine the percentage of the population that has a certain characteristic, holds a particular opinion, or engages in a particular behavior.

Differential carryover The response to one treatment depends on which treatment was administered previously.

Diffusion or imitation of treatment A threat to internal validity that can occur if participants in one treatment group become familiar with the treatment of another group and copy that treatment.

Directional research hypothesis Prediction of the specific outcome of an experiment.

Discussion section The fourth major section of the APA-format paper. Contains a summary of the experiment's results, a comparison of those results to previous research, and the conclusion(s) drawn from the experiment.

Double-blind experiment An experiment in which both the experimenter and the participants

are unaware of which treatment the participants are receiving.

Effect size The magnitude or size of the experimental treatment.

Elimination A control technique whereby extraneous variables are completely removed from an experiment.

Emic A culture-specific finding.

Empirical Objectively quantifiable observations.

Environmental generalization Applying the results from an experiment to a situation or environment that differs from that of the original experiment.

Error variability Variability in DV scores that is due to factors other than the IV, such as individual differences, measurement error, and extraneous variation (also known as *within-groups variability*).

Ethnocentric Other cultures are viewed as an extension of one's own culture.

Ethnographic inquiry Research in which the goal is to learn about a culture or some aspect of a culture from the perspective of the members of that culture.

Etic A finding that is the same in different cultures.

Experience IV Manipulation of the amount or type of training or learning.

Experiment An attempt to determine the cause-and-effect relations that exist in nature. Involves the manipulation of an independent variable (IV), recording of changes in a dependent variable (DV), and control of extraneous variables.

Experimental analysis of behavior A research approach popularized by B. F. Skinner, in which a single participant is studied.

Experimental design The general plan for selecting participants, assigning participants to experimental conditions, controlling extraneous variables, and gathering data.

Experimental group In a two-group design, the group of participants that receives the IV.

Experimental mortality This threat to internal validity can occur if experimental participants from different groups drop out of the experiment at different rates.

Ex post facto research A research approach in which the experimenter cannot directly manipulate the IV but can only classify, categorize, or measure the IV because it is predetermined in the participants (e.g., IV = sex).

Ex post facto study A study in which the variable(s) to be studied are selected after they have occurred.

External validity A type of evaluation of an experiment; do the experimental results apply to populations and situations that are different from those of the experiment?

Extraneous variables Uncontrolled variables that may unintentionally influence the dependent variable (DV) and thus invalidate an experiment.

Fabrication of data Those instances where the experimenter either deliberately alters or creates research data.

Factorial design An experimental design with more than one IV.

Factors Synonymous with IVs.

Figure A pictorial representation of a set of results.

Focus group Seven to 10 participants with shared experiences or similar characteristics who meet for 1 to 1½ hours to discuss a topic of common interest.

Frequency polygon A graph that is constructed by placing a dot in the center of each bar of a histogram and then connecting the dots.

General implication form Statement of the research hypothesis in an "if . . . then" form.

Generalization Applying the results from an experiment to a different situation or population.

Good participant effect The tendency of participants to behave as they perceive the experimenter wants them to behave.

Grounded theory A qualitative research approach that attempts to develop theories of understanding based on data from the real world.

Hawthorne effect Another name for reactivity or reactivity effect.

Headings Titles for various sections of a psychology paper that are designed to help the reader understand the outline and importance of the parts of the paper.

Heterogeneity of variance Occurs when we do not have homogeneity of variance; this means that our two (or more) groups' variances are not equivalent.

Histogram A graph in which the frequency for each category of a quantitative variable is represented as a vertical column that touches the adjacent column.

History A threat to internal validity; refers to events that occur between the DV measurements in a repeated-measures design.

Homogeneity of variance The assumption that the variances are equal for the two (or more) groups you plan to compare statistically.

Hypothesis An attempt to organize certain data and specific relations among variables within a specific portion of a larger, more comprehensive theory.

Incomplete counterbalancing Only a portion of all possible sequences are presented.

Independent groups Groups of participants formed by random assignment.

Independent variable (IV) A stimulus or aspect of the environment that the experimenter directly manipulates to determine its influences on behavior.

Inductive logic Reasoning that proceeds from specific cases to general conclusions or theories.

Inferential statistics Procedures used to analyze data after an experiment is completed in order to determine whether the independent variable has a significant effect.

Institutional Review Board (IRB) The university committee that is responsible for determining whether a proposed research project conforms to accepted ethical standards.

Instrumentation A threat to internal validity that occurs if the equipment or human measuring the DV changes the measuring criterion over time.

Interaction The joint, simultaneous effect on the DV of more than one IV.

Interaction of selection and treatment A threat to external validity that can occur when a treatment effect is found only for a specific sample of participants.

Interaction of testing and treatment A threat to external validity that occurs when a pretest sensitizes participants to the treatment yet to come.

Interactions with selection Threats to internal validity that can occur if there are systematic differences between or among selected treatment groups based on maturation, history, or instrumentation.

Internal validity A type of evaluation of your experiment; it asks whether your IV is the only possible explanation of the results shown for your DV.

Interobserver reliability The extent to which observers agree.

Interrater reliability Degree of agreement among judges concerning the content validity of test or inventory items.

Interrupted time-series design A quasi-experimental design, involving a single group of participants, that includes repeated pretreatment measures, an applied treatment, and repeated posttreatment measures.

Interval scale A scale of measurement that permits rank ordering of events with the assumption of equal intervals between adjacent events.

Introduction The first major section of the APA-format paper. Contains the thesis statement, review of relevant literature, and experimental hypothesis.

Level 1 heading A centered section title in which the first letters of major words are capitalized. Occupies a line by itself.

Level 3 heading A section title that is left margin justified, italicized, and has the first letter of each major word capitalized. Occupies a line by itself.

Level 4 heading A section title that is indented five spaces, italicized, has only the first word capitalized, and ends with a period; it does not occupy a separate line.

Levels Differing amounts or types of an IV used in an experiment (also known as *treatment conditions*).

Line graph A graph that is frequently used to depict the results of an experiment.

Longitudinal research project Obtaining research data from the same group of participants over an extended period of time.

Main effect Refers to the sole effect of one IV in a factorial design.

Manuscript page header The first two or three words of the report's title. Appears with the page number on each page of the research report.

Marginal significance Refers to statistical results with a probability of chance between 5% and 10%; in other words, almost significant, but not quite. Researchers often talk about such results as if they reached the $p = .05$ level.

Matched pairs Research participants in a two-group design who are measured and equated on some variable before the experiment.

Matching variable A potential extraneous variable on which we measure our research participants and from which we form sets of participants who are equal on the variable.

Materials subsection The second subsection of the method section. When appropriate, contains information about materials other than the equipment used in the experiment.

Maturation A threat to internal validity; refers to changes in participants that occur over time during an experiment; could include actual

physical maturation or tiredness, boredom, hunger, and so on.

Measurement The assignment of symbols to events according to a set of rules.

Mean The arithmetic average of a set of numbers; found by adding all the scores in a set and then dividing by the number of scores.

Mean square The “averaged” variability for each source; computed by dividing each source’s sum of squares by its degrees of freedom.

Median The number that divides a distribution in half.

Method section The second major section of the APA-format paper. Contains information about the participants; the apparatus, materials, and testing instrument(s), and the procedures used in the experiment.

Mixed assignment A factorial design that has a mixture of independent groups for one IV and correlated groups for another IV. In larger factorial designs at least one IV has independent groups and at least one has correlated groups (also known as *mixed groups*).

Mode The score in a distribution that occurs most often.

Mortality A threat to internal validity that can occur if experimental participants from different groups drop out of the experiment at different rates.

Multiple-treatment interference A threat to external validity that occurs when a set of findings results only when participants experience multiple treatments in the same experiment.

Natural pairs Research participants in a two-group design who are naturally related in some way (e.g., a biological or social relationship).

Naturalistic observation Seeking answers to research questions by observing behavior in the real world.

Nay-sayers Participants who tend to answer no to all questions.

Negative correlation As scores on one variable increase, scores on the second variable decrease.

Nominal scale A scale of measurement in which events are assigned to categories.

Nondirectional research hypothesis A specific prediction concerning the outcome of an experiment is not made.

Nonequivalent group design A design involving two or more groups that are not randomly assigned; a comparison group (no treatment) is compared to one or more treatment groups.

Nonreactive measures DV measurements that do not influence the DV being measured.

Nonsystematic sources Sources for research ideas that present themselves in an unpredictable manner; a concerted attempt to locate researchable ideas has not been made.

Normal distribution A symmetrical, bell-shaped distribution having half the scores above the mean and half the scores below the mean.

Nuisance variable Unwanted variables that can cause the variability of scores within groups to increase.

Null hypothesis A hypothesis that says that all differences between groups are due to chance (i.e., not the operation of the IV).

One-way ANOVA A statistical test used to analyze data from an experimental design with one independent variable that has three or more groups (levels).

Open coding The process of describing data through means such as examination, comparison, conceptualization, and categorization.

Operational definition Defining the independent, dependent, and extraneous variables in terms of the operations needed to produce them.

Ordinal scale A scale of measurement that permits events to be rank ordered.

Ordinate The vertical or y axis of a graph.

Participant characteristics Aspects of the participant, such as age, sex, or personality traits, which are treated as if they were IVs.

Participant observation Research in which the researcher becomes part of the group being studied.

Participants at minimal risk Participants in an experiment that does not place them under physical or emotional risk.

Participants at risk Participants in an experiment that places them under some type of physical or emotional risk.

Participants subsection The first subsection of the method section. Provides full information about the participants in the study.

Personality test or inventory Measures a specific aspect of the individual's motivational state, interpersonal capability, or personality.

Physiological IV A physiological state of the participant manipulated by the experimenter.

Pie chart Graphical representation of the percentage allocated to each alternative as a slice of a circular pie.

Placebo effect An experimental effect caused by expectation or suggestion rather than the IV.

Plagiarism Using someone else's work without giving credit to the original source.

Pilot testing Preliminary, exploratory testing that is done prior to the complete research project.

Population The complete set of individuals or events.

Population generalization Applying the results from an experiment to a group of participants that is different and more encompassing than those used in the original experiment.

Positive correlation As scores on one variable increase, scores on the second variable also increase.

Post hoc comparisons Statistical comparisons made between group means after finding a significant F ratio.

Power The probability that a statistical test will be significant (i.e., the experimental hypothesis is accepted when it is true).

Practice effect A beneficial effect on a DV measurement caused by previous experience with the DV.

Precedent An established pattern.

Principle of falsifiability Results not in accord with the research hypothesis are taken as evidence that this hypothesis is false.

Principle of parsimony The belief that explanations of phenomena and events should remain simple until the simple explanations are no longer valid.

Procedure subsection The third subsection of the method section. Provides a step-by-step account of what the participants and experimenter did during the experiment.

Process The manner in which actions and interactions occur in a sequence or series.

Programmatic research A series of research experiments concerning a related topic or question.

Qualitative research Research conducted in a natural setting that seeks to understand a complex human behavior by developing a complete narrative description of that behavior.

Quasi-experimental design A research design used when the researcher cannot randomly assign experimental participants to the groups but the researcher does manipulate an IV and measure a DV.

Random assignment A method of assigning research participants to groups so that each participant has an equal chance of being in any group.

Random sample A sample in which every member of the population has an equal likelihood of being included.

Random sampling without replacement Once chosen, a score, event, or participant cannot be returned to the population to be selected again.

Random sampling with replacement Once chosen, a score, event, or participant can be returned to the population to be selected again.

Random selection A control technique that ensures that each member of the population has an equal chance of being chosen for an experiment.

Randomization A control technique that ensures that each participant has an equal chance of being assigned to any group in an experiment.

Range A measure of variability that is computed by subtracting the smallest score from the largest score.

Ratio scale A scale of measurement that permits rank ordering of events with the assumptions of equal intervals between adjacent events and a true zero point.

Reactance or reactivity effect The finding that participants respond differently when they know they are being observed.

Reactive arrangements A threat to external validity caused by an experimental situation that alters participants' behavior, regardless of the IV involved.

Reactive measures DV measurements that actually change the DV being measured.

Reference A full bibliographic record of any work cited in the text of a psychological paper.

Reference section A complete listing of all the references cited in a psychological paper.

Reliable Producing consistent measurements.

Reliability Extent to which a test or inventory is consistent in its evaluation of the same individuals.

Repeated measures An experimental procedure in which research participants are tested or measured more than once.

Repeated-measures ANOVA This one-way ANOVA uses correlated groups of participants.

Replication An additional scientific study that is conducted in exactly the same manner as the original research project.

Replication with extension An experiment that seeks to confirm (replicate) a previous finding but does so in a different setting or with different participants or under different conditions.

Research design The general plan for conducting research and gathering data.

Research idea Identification of a gap in the knowledge base or an unanswered question in an area of interest.

Research or experimental hypothesis The experimenter's predicted outcome of a research project.

Response set The result when an experimental context or testing situation influences the participants' responses.

Results section The third major section of the APA-format paper. Contains information about the statistical findings from the experiment.

Robust Refers to a statistical test that can tolerate violation of its assumptions (e.g., homogeneity of variances) and still yield valid results.

Rosenthal effect The result when an experimenter's preconceived idea of appropriate responding influences the treatment of participants and their behavior.

Running head A condensed title that is printed at the top of alternate pages of a published article.

Sample A group that is selected to represent the population.

Scale of measurement A set of measurement rules.

Selection A threat to internal validity that can occur if participants are chosen in such a way that the groups are not equal before the experiment; the researcher cannot then be certain that the IV caused any difference observed after the experiment.

Selective coding The process of selecting the main phenomenon (core category) around which all other phenomena (subsidiary categories) are grouped, arranging the groupings, studying the results, and rearranging where necessary.

Sequence or order effects The position of a treatment in a series determines, in part, the participants' response.

Serendipity A situation in which one phenomenon is sought but something else is found.

Single-blind experiment An experiment in which the experimenter (or participants) is unaware of the treatment the participants are receiving.

Single-case experimental design An experiment that consists of one participant (also known as $N = 1$ designs).

Single-strata approach Gathering data from a single stratum of the population of interest.

Situation sampling Observing the same behavior in different situations.

Source table A table that contains the results of ANOVA. Source refers to the source of the different types of variation.

Split-half technique Determination of reliability by dividing the test or inventory into two subtests and then comparing the scores made on the two halves.

Standard deviation Square root of the variance; has important relations to the normal curve.

Statistical regression A threat to internal validity that occurs when low scorers improve or high scorers fall on a second administration of a test solely as a result of statistical reasons.

Statistics The branch of mathematics that involves the collection, analysis, and interpretation of data.

Stimulus or environmental IV An aspect of the environment manipulated by the experimenter.

Stratified random sampling Random samples are drawn from specific subpopulations or strata of the general population.

Sum of squares The amount of variability in the DV attributable to each source.

Synergistic effects Dramatic consequences that occur when you combine two or more substances, conditions, or organisms. The effects are greater (or less) than what is individually possible.

Synthetic statement Statements that can be either true or false.

Systematic sources Thoroughly examined, carefully thought-out sources for research topics.

t test An inferential statistical test used to evaluate the difference between two means.

Table A chart containing an array of descriptive statistics.

Temporal generalization Applying the results from an experiment to a time that is different from the time when the original experiment was conducted.

Testing A threat to internal validity that occurs because measuring the DV causes a change in the DV.

Testing instrument(s) subsection The second subsection of the method section. When appropriate, contains information about standardized tests used in the experiment.

Test-retest procedure Determination of reliability by repeatedly administering a test to the same participants.

Theory A formal statement of the relations among the IVs and DVs in a given area of research.

Thesis statement A statement of the general research topic of interest and the perceived relation of the relevant variables in that area.

Three-way design A factorial design with three IVs.

Time sampling Making observations at different time periods.

Title page The first page of an APA-format paper. It includes the manuscript page header, the running head, the manuscript's title, and the name(s) of the author(s) and their affiliation(s).

Transactional system An analysis of how actions and interactions relate to their conditions and consequences.

Transferability The extent to which the results of a qualitative research project can be generalized to other settings and groups.

Treatment groups Groups of participants that receive the IV.

Treatment variability Variability in DV scores due to the effects of the IV (also known as *between-groups variability*).

True experiment An experiment in which the experimenter directly manipulates the IV.

Type I error Accepting the experimental hypothesis when the null hypothesis is true.

Type II error Accepting the null hypothesis when the experimental hypothesis is true.

Unbiased language Language that does not display prejudice toward an individual or group.

Valid Measuring what is supposed to be measured.

Validity The extent to which a test or inventory measures what it is supposed to measure.

Variability The extent to which scores spread out around the mean.

Variable An event or behavior that can assume two or more values.

Variance A single number that represents the total amount of variation in a distribution; also the square of the standard deviation, σ^2 .

Within-group counterbalancing Presentation of different treatment sequences to different participants.

Within-groups variability Another term for error variability.

Within-subject counterbalancing Presentation of different treatment sequences to the same participant.

Within-subjects comparison Refers to a contrast between groups of participants who were assigned to groups through matched pairs, natural pairs, or repeated measures.

Yea-sayers Participants who tend to answer yes to all questions.

Zero correlation Two variables under consideration are not related.

References

- Adair, J. G., & Vohra, N. (2003). The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. *American Psychologist*, 58, 15–23.
- The American Enterprise. (1993, September/October). p. 101. From Figure 9-3.
- American Psychological Association. (1973). *Ethical principles in the conduct of research with human participants*. Washington, DC: Author.
- American Psychological Association. (1982). *Ethical principles in the conduct of research with human participants*. Washington, DC: Author.
- American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- American Psychological Association. (2001). *Thesaurus of psychological index terms* (9th ed.). Washington, DC: Author.
- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060–1073.
- American Psychological Association. (2003). *APA Style.org: Electronic references*. Retrieved July 17, 2005, from <http://www.apastyle.org/elecref.html>
- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Aronson, E., & Carlsmith, J. M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (2nd ed., pp. 1–79). Reading, MA: Addison-Wesley.
- Ball, R. M., Kargl, E. S., Kimpel, J. D., & Siewert, S. L. (2001). Effect of sad and suspenseful mood induction on reaction time. *Psi Chi Journal of Undergraduate Research*, 6, 21–27.
- Barlow, D. H., & Hersen, M. (1973). Single-case experimental designs. *Archives of General Psychiatry*, 29, 319–325.
- Baumrind, D. (1964). Some thoughts on the ethics of research: After reading Milgram's "Behavioral study of obedience." *American Psychologist*, 19, 421–423.
- Beach, F. A. (1950). The snark was a boojum. *American Psychologist*, 5, 115–124.
- Beauchamp, T. L., & Childress, J. F. (1979). *Principles of biomedical ethics*. New York: Oxford University Press.
- Bellquist, J. E. (1993). *A guide to grammar and usage for psychology and related fields*. Hillsdale, NJ: Erlbaum.
- Benjamin, L. T., Jr. (1992, April). *Health, happiness, and success: The popular psychology magazine of the 1920s*. Paper presented at the annual meeting of the Eastern Psychological Association, Boston, MA.
- Bennett, S. K. (1994). The American Indian: A psychological overview. In W. J. Lonner & R. Malpass (Eds.), *Psychology and culture* (pp. 35–39). Boston: Allyn and Bacon.
- Blumenthal, A. L. (1991). The intrepid Joseph Jastrow. In G. A. Kimble, M. Wertheimer, & C. L. White (Eds.), *Portraits of pioneers in*

- psychology* (pp. 75–87). Washington, DC: American Psychological Association.
- Bodner, D., Cochran, C. D., & Blum, T. L. (2000). Unique invulnerability measurement in skydivers: Scale validation. *Psi Chi Journal of Undergraduate Research*, 5, 104–108.
- Boice, M. L., & Gargano, G. J. (2001). Part-set cuing is due to strong, not weak, list cues. *Psi Chi Journal of Undergraduate Research*, 6, 118–122.
- Boring, E. G. (1929). *A history of experimental psychology*. New York: The Century Co.
- Brewer, C. L. (2002). Reflections on an academic career: From which side of the looking glass? In S. F. Davis & W. Buskist, *The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer* (pp. 499–507). Mahwah, NJ: Erlbaum.
- Bridgman, P. W. (1927). *The logic of modern physics*. New York: Macmillan.
- Broad, W. J. (1980). Imbroglio at Yale (I): Emergence of a fraud. *Science*, 210, 38–41.
- Burkley, E., III, McFarland, S., Walker, W., & Young, J. (2000). Informed consent: A methodological confound? *Psi Chi Journal of Undergraduate Research*, 5, 43–45.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54, 297–312.
- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, 24, 409–429.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.
- Claus, M. K. (2000). Gendered toy preferences and assertiveness in preschoolers. *Journal of Psychological Inquiry*, 5, 15–18.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.
- Corkin, S. (1984). Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings in H. M. *Seminars in Neurology*, 4, 249–259.
- Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Darley, J. M., & Latané, D. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377–383.
- Davis, S. F., Grover, C. A., & Erickson, C. A. (1987). A comparison of the aversiveness of denatonium saccharide and quinine in humans. *Bulletin of the Psychonomic Society*, 25, 462–463.
- Davis, S. F., Grover, C. A., Erickson, C. A., Miller, L. A., & Bowman, J. A. (1987). Analyzing the aversiveness of denatonium saccharide. *Perceptual and Motor Skills*, 64, 1215–1222.
- Department of Psychology, Bishop's University. *Plagiarism pamphlet*. Supplied December 23, 1994.
- Dickson, A. J., Giuliano, T. A., Morris, J. C., & Cass, K. L. (2001). Eminem versus Charley Pride: Race, stereotypes, and perceptions of rap and country music performers. *Psi Chi Journal of Undergraduate Research*, 6, 175–180.
- Doyle, A. C. (1927). *The complete Sherlock Holmes*. Garden City, NY: Doubleday.
- Driggers, K. J., & Helms, T. (2000). The effects of salary on willingness to date. *Psi Chi Journal of Undergraduate Research*, 5, 76–80.
- Dukes, W. F. (1965) *N = 1*. *Psychological Bulletin*, 64, 74–79.
- Dunham, R. G., & Yandell, L. R. (2005). A measure of drawing self-efficacy: A psychometric evaluation. *Psi Chi Journal of Undergraduate Research*, 10, 16–20.
- Dunn, J., Ford, K., Rewey, K. L., Juve, J. A., Weiser, A., & Davis, S. F. (2001). A modified presubmission checklist. *Psi Chi Journal of Undergraduate Research*, 6, 142–144.

- Early, K. E., Holloway, A. L., & Welton, G. L. (2005). A look into forgiveness: The roles of intentionality, severity, and friendship. *Psi Chi Journal of Undergraduate Research*, 10, 54–59.
- Featherston, J. (2008). Qualitative research. In S. F. Davis & W. Buskist (Eds.), *21st century psychology: A reference handbook* (pp. 93–100). Los Angeles: Sage.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Fine, M., Torre, M. E., Boudin, K., Bowen, I., Clark, J., Hylton, D., et al. (2003). Participatory action research: From within and beyond prison bars. In P. M. Camic, J. E. Rhodes, & L. Yardley (Eds.), *Qualitative research in psychology: Empirical studies of scientific change* (pp. 197–230). Baltimore: Johns Hopkins University Press.
- Franklin, M. B. (1990). Reshaping psychology at Clark: The Werner Era. *Journal of the History of the Behavioral Sciences*, 26, 176–189.
- Freeman, N. J., & Punzo, D. L. D. (2001). Effects of DNA and eyewitness evidence on juror decisions. *Psi Chi Journal of Undergraduate Research*, 6, 109–117.
- Gauthraux, V. M. (2000). Relating pretrial publicity, judicial instruction, and thought suppression with guilt ratings. *Psi Chi Journal of Undergraduate Research*, 5, 21–28.
- Geronimus, A. T. (1991). Teenage childbearing and social and reproductive disadvantage: The evolution of complex questions and the demise of simple answers. *Family Relations*, 40, 463–471.
- Gibson, A. R., Smith, K., & Torres, A. (2000). Glancing behavior of participants using automated teller machines while bystanders approach. *Psi Chi Journal of Undergraduate Research*, 5, 148–151.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117, 21–38.
- Glesne, C. (1999). *Becoming qualitative researchers: An introduction* (2nd ed.). New York: Longman.
- Goodwin, C. J. (2005). *A history of modern psychology* (2nd ed.). Hoboken, NJ: Wiley.
- Gray, A. D., & Lucas, J. L. (2001). Commuters' subjective perceptions of travel impedance and their stress levels. *Psi Chi Journal of Undergraduate Research*, 6, 79–83.
- Grazer, B. (Producer), & Howard, R. (Producer/Director). 2001. *A beautiful mind* [Motion picture]. United States: Universal Studios.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 60–69). Thousand Oaks, CA: Sage.
- Guerin, B. (1986). Mere presence effects in humans: A review. *Journal of Personality and Social Psychology*, 22, 38–77.
- Guthrie, R. V. (1976). *Even the rat was white: A historical view of psychology*. New York: Harper & Row.
- Hall, R. V., Fox, R., Willard, D., Goldsmith, L., Emerson, M., Owen, M., et al. (1971). The teacher as observer and experimenter in the modification of disputing and talking-out behaviors. *Journal of Applied Behavior Analysis*, 4, 141–149.
- Hatch, J. A. (2002). *Doing qualitative research in educational settings*. Albany: State University of New York Press.
- Hayes, K. M., Miller, H. R., & Davis, S. F. (1993). *Examining the relationships between interpersonal flexibility, self-esteem, and death anxiety*. Paper presented at the annual meeting of the Southwestern Psychological Association, Corpus Christi, TX.
- Hedrick, T. E., Bickman, L., & Rog, D. J. (1993). *Applied research design: A practical guide*. Newbury Park, CA: Sage.
- Hedrick, T. E., & Shipman, S. L. (1988). Multiple questions require multiple designs: An evaluation of the 1981 changes to the AFDC program. *Evaluation Review*, 12, 427–448.

- Helmreich, R., & Stapp, J. (1974). Short form of the Texas Social Behavior Inventory (TSBI), an objective measure of self-esteem. *Bulletin of the Psychonomic Society*, 4, 473–475.
- Hersen, M. (1982). Single-case experimental designs. In A. S. Bellack, M. Hersen, & A. E. Kazdin (Eds.), *International handbook of behavior modification and therapy* (pp. 167–203). New York: Plenum Press.
- Hersen, M., & Barlow, D. H. (1976). *Single-case experimental designs: Strategies for studying behavioral change*. New York: Pergamon Press.
- Hilts, P. J. (1995). *Memory's ghost: The strange tale of Mr. M and the nature of memory*. New York: Simon & Schuster.
- Horvat, J., & Davis, S. F. (1998). *Doing psychological research*. Upper Saddle River, NJ: Prentice Hall.
- Howard, D. (1997). Language in the human brain. In M. D. Rugg (Ed.), *Cognitive neuroscience* (pp. 277–304). Cambridge, MA: MIT Press.
- Huff, D. (1954). *How to lie with statistics*. New York: Norton.
- Ittayem, N. M., & Cooley, E. L. (2004). Self-disclosure of emotional experiences: Narrative writing and drawing for stress reduction. *Psi Chi Journal of Undergraduate Research*, 9, 126–133.
- James, N., & Pritchard, M. (2005). The relationship between disordered eating and stress, class year, and figure dissatisfaction among college students. *Psi Chi Journal of Undergraduate Research*, 10, 60–65.
- Jones, J. H. (1981). *Bad blood: The Tuskegee syphilis experiment*. New York: Free Press.
- Jones, J. M. (1994). The African American: A duality dilemma? In W. J. Lonner & R. Malpass (Eds.), *Psychology and culture* (pp. 17–21). Boston: Allyn and Bacon.
- Jones, M. D. (2001). The effects of noise and sex on children's performance on recall and spatial tasks. *Psi Chi Journal of Undergraduate Research*, 6, 63–67.
- Jones, R. R., Vaught, R. S., & Weinrott, M. (1977). Time-series analysis in operant research. *Journal of Applied Behavior Analysis*, 10, 151–166.
- Kalat, J. (2001). *Biological psychology* (7th ed.). Belmont, CA: Wadsworth.
- Kaser, R. T. (1998). Secondary information services—Mirrors of scholarly communication: Forces and trends. In A. Henderson (Ed.), *Electronic databases and publishing* (pp. 9–23). New Brunswick, NJ: Transaction.
- Kazdin, A. E. (1976). Statistical analyses for single-case experimental designs. In M. Hersen & D. H. Barlow (Eds.), *Single-case experimental designs: Strategies for studying behavioral change* (pp. 265–316). New York: Pergamon Press.
- Kazdin, A. E. (1984). *Behavior modification in applied settings* (3rd ed.). Homewood, IL: Dorsey Press.
- Kazlauskas, M. A., & Kelland, M. D. (1997). Behavioral evaluation of the stargazer mutant rat in a tactile startle paradigm. *Psi Chi Journal of Undergraduate Research*, 2, 90–98.
- Keith-Spiegel, P., & Wiederman, M. W. (2000). *The complete guide to graduate school admission: Psychology, counseling, and related professions* (2nd ed.). Mahwah, NJ: Erlbaum.
- Keppel, G., Saufley, W. H., Jr., & Tokunaga, H. (1992). *Introduction to design & analysis: A student's handbook* (2nd ed.). New York: W. H. Freeman.
- Kiker, K. (2001). Developmental changes in children's measurement errors. *Psi Chi Journal of Undergraduate Research*, 6, 68–74.
- Kimmel, A. J. (1988). *Ethics and values in applied social research*. Beverly Hills, CA: Sage.
- Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Brooks/Cole.
- Kirk, R. E. (1996, April). *Practical significance: A concept whose time has come*. Southwestern Psychological Association presidential address. Houston, TX.

- Koestler, A. (1964). *The act of creation*. New York: Macmillan.
- Kohn, P. M., Lafreniere, K., & Gurevich, M. (1990). The Inventory of College Students' Recent Life Experiences: A decontaminated hassles scale for a special population. *Journal of Behavioral Medicine*, 13, 619–630.
- Korn, J. H. (1988). Students' roles, rights, and responsibilities as research participants. *Teaching of Psychology*, 15, 74–78.
- Krantz, D. S., Glass, D. C., & Snyder, M. L. (1974). Helplessness, stress level, and the coronary-prone behavior pattern. *Journal of Experimental Social Psychology*, 10, 284–300.
- Landrum, R. E., & Davis, S. F. (2006). *The psychology major: Career options and strategies for success* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Lang, P. J., & Melamed, B. G. (1969). Avoidance conditioning therapy of an infant with chronic ruminative vomiting. *Journal of Abnormal Psychology*, 74, 1–8.
- Langley, W. M., Theis, J., Davis, S. F., Richard, M. M., & Grover, C. A. (1987). Effects of denatonium saccharide on the drinking behavior of the grasshopper mouse (*Onychomys leucogaster*). *Bulletin of the Psychonomic Society*, 25, 17–19.
- Larey, A. (2001). Effect of massage and informal touch on body-dissatisfaction, anxiety, and obsessive-compulsiveness. *Journal of Psychological Inquiry*, 6, 78–83.
- Lee, D. J., & Hall, C. C. I. (1994). Being Asian in North America. In W. J. Lonner & R. Malpass (Eds.), *Psychology and culture* (pp. 29–33). Boston: Allyn and Bacon.
- Licht, C. A. (2000). Occupational stress as a function of type of organization and sex of employee. *Psi Chi Journal of Undergraduate Research*, 5, 46–55.
- Licht, C. A., & Solomon, L. Z. (2001). Occupational stress as a function of type of organization and sex of employee: A reassessment. *Psi Chi Journal of Undergraduate Research*, 6, 14–20.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Lonner, W. J., & Malpass, R. (1994a). Preface. In W. J. Lonner & R. Malpass (Eds.), *Psychology and culture* (pp. ix–xiii). Boston: Allyn and Bacon.
- Lonner, W. J., & Malpass, R. (1994b). When psychology and culture meet: An introduction to cross-cultural psychology. In W. J. Lonner & R. Malpass (Eds.), *Psychology and culture* (pp. 1–12). Boston: Allyn and Bacon.
- Losch, M. E., & Cacioppo, J. T. (1990). Cognitive dissonance may enhance sympathetic tonus, but attitudes are changed to reduce negative affect rather than arousal. *Journal of Experimental Social Psychology*, 26, 289–304.
- Lucas, G. M., Rainville, B., Bhan, P., Rosenberg, J., Proud, K., & Koger, S. M. (2005). Memory for computer-generated graphics: Boundary extension in photographic vs. computer-generated images. *Psi Chi Journal of Undergraduate Research*, 10, 43–48.
- Luehring, J., & Altman, J. D. (2000). Factors contributing to sex differences on the mental rotation task. *Psi Chi Journal of Undergraduate Research*, 5, 29–35.
- Mahoney, M. J. (1987). Scientific publication and knowledge politics. *Journal of Social Behavior and Personality*, 2, 165–176.
- Marín, G. (1994). The experience of being a Hispanic in the United States. In W. J. Lonner & R. Malpass (Eds.), *Psychology and culture* (pp. 23–27). Boston: Allyn and Bacon.
- Marzola, L. M., & Brooks, C. I. (2004). Journal citation frequency in the *Psi Chi Journal of Undergraduate Research*. *Psi Chi Journal of Undergraduate Research*, 9, 146–149.
- Matsumoto, D. (2000). *Culture and psychology: People around the world* (2nd ed.). Belmont, CA: Wadsworth.
- McClellan, K. S., & Woods, E. B. (2001). Disability and society: Appearance stigma results in discrimination toward deaf persons. *Psi*

- Chi Journal of Undergraduate Research*, 6, 57–62.
- McGuigan, F. J. (1960). *Experimental psychology*. Englewood Cliffs, NJ: Prentice Hall.
- McInerny, R. M. (1970). *A history of Western philosophy*. Notre Dame, IN: University of Notre Dame Press.
- McKibban, A. R., & Nelson, S. (2001). Subjective well-being: Are you busy enough to be happy? *Psi Chi Journal of Undergraduate Research*, 6, 51–56.
- Medewar, P. B. (1979). *Advice to a young scientist*. New York: Harper & Row.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67, 371–378.
- Milgram, S. (1964). Issues in the study of obedience: A reply to Baumrind. *American Psychologist*, 19, 848–852.
- Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.
- Miller, H. B., & Williams, W. H. (1983). *Ethics and animals*. Clifton, NJ: Humana Press.
- Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, 46, 961–978.
- Miller, N. E. (1985). The value of behavioral research on animals. *American Psychologist*, 40, 423–440.
- Montgomery, M. D., & Donovan, K. M. (2002, April). The “Mozart effect” on cognitive and spatial task performance, revisited. Poster presented at the annual meeting of the Southwestern Psychological Association, Corpus Christi, TX.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38, 379–387.
- Mooney, R. L. (1950). *Mooney problem check list*. New York: The Psychological Corporation.
- Nash, S. M. (1983). *The relationship between early ethanol exposure and adult taste preference*. Paper presented in J. P. Guilford/Psi Chi undergraduate research competition.
- Novaco, R. W., Stokols, D., & Milanesi, L. (1990). Objective and subjective dimensions of travel impedance as determinants of commuting stress. *American Journal of Community Psychology*, 18, 231–257.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776–783.
- Pepperberg, I. M. (1994). Numerical competence in an African gray parrot (*Psittacus erithacus*). *Journal of Comparative Psychology*, 108, 36–44.
- Peshkin, A. (1986). *God's choice: The total world of a fundamentalist Christian school*. Chicago: University of Chicago Press.
- Prieto, L. R. (2005). The IRB and psychological research: A primer for students. *Eye on Psi Chi*, 9, 24–25.
- Proctor, R. W., & Capaldi, E. J. (2001). Improving the science education of psychology students: Better teaching of methodology. *Teaching of Psychology*, 28, 173–181.
- Reagan, T. (1983). *The case for animal rights*. Berkeley: University of California Press.
- Reid, D. W., & Ware, E. E. (1973). Multidimensionality of internal-external control: Implications for past and future research. *Canadian Journal of Behavioural Research*, 5, 264–271.
- Riepe, S. M. (2004). Effects of education level and gender on achievement motivation. *Psi Chi Journal of Undergraduate Research*, 9, 33–38.
- Robinson, K. (2005). Effects of environmental factors on the health of college students. *Psi Chi Journal of Undergraduate Research*, 10, 3–8.
- Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker*. Cambridge, MA: Harvard University Press.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.

- Rosenthal, R. (1977). Biasing effects of experimenters. *ETC: A review of general semantics*, 34, 253–264.
- Rosenthal, R. (1985). From unconscious experimenter bias to teacher expectancy effects. In J. B. Dusek, V. C. Hall, & W. J. Meyer (Eds.), *Teacher expectancies* (pp. 37–65). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, 8, 183–189.
- Rosenthal, R., & Jacobson, L. (1968). *Psychodynamics in the classroom*. New York: Holt.
- Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research*. New York: McGraw-Hill.
- Rosenthal, R., & Rosnow, R. (1991). *Essentials of behavioral research* (2nd ed.). New York: McGraw-Hill.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York: Academic Press.
- Sagles, S. E., Coley, S., Espiritu, G., Zahregian, P., & Velayo, R. (2002). Effects of racial background and sex on identifying facial expressions and person perception. *Psi Chi Journal of Undergraduate Research*, 7, 31–37.
- Salhany, J., & Roig, M. (2004). Academic dishonesty policies across universities: Focus on plagiarism. *Psi Chi Journal of Undergraduate Research*, 9, 150–153.
- Sasson, R., & Nelson, T. M. (1969). The human experimental subject in context. *Canadian Psychologist*, 10, 409–437.
- Scali, R. M., & Brownlow, S. (2001). Impact of instructional manipulation and stereotype activation on sex differences in spatial task performance. *The Psi Chi Journal of Undergraduate Research*, 6, 3–13.
- Schroepel, N., & Carlo, G. (2001). Sympathy and perspective taking as mediators of gender differences in altruism. *Journal of Psychological Inquiry*, 6, 7–11.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51, 515–530.
- Serrano-Rodriguez, J., Brunolli, S., & Echolds, L. (2007). Approval of religious attitudes on approval of organ donation. *Journal of Psychological Inquiry*, 12, 63–68.
- Shuter, R. (1977). A field study of nonverbal communication in Germany, Italy, and the United States. *Communication Monographs*, 44, 298–305.
- Singer, P. (1975). *Animal liberation*. New York: Random House.
- Skinner, B. F. (1961). *Cumulative record* (expanded edition). New York: Appleton-Century-Crofts.
- Skinner, B. F. (1966). Operant behavior. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 12–32). New York: Appleton-Century-Crofts.
- Small, W. S. (1901). An experimental study of the mental processes of the rat. *American Journal of Psychology*, 11, 133–165.
- Smith, L. N., & Gaines, M. T. IV. (2005). Prescription medication, backward masking performance, and attention deficit hyperactivity disorder. *Psi Chi Journal of Undergraduate Research*, 10, 9–15.
- Smith, R. A. (1998). Another perspective on publishing: Keeping the editor happy. *Psi Chi Journal of Undergraduate Research*, 3, 51–55.
- Smith, R. A., Davis, S. F., & Burleson, J. (1996). *Snark hunting revisited: The status of comparative psychology in the 1990s*. Paper presented at the meeting of the Southwestern Comparative Psychology Association, Houston, TX.

- Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin, 46*, 137–150.
- Sou, E. K., & Irving, L. M. (2002). Students' attitudes toward mental illness: A Macao-U.S. cross-cultural comparison. *Psi Chi Journal of Undergraduate Research, 7*, 13–22.
- Spatz, C. (2001). *Basic statistics: Tales of distributions* (7th ed.). Belmont, CA: Wadsworth.
- Steele, C. M., Southwick, L. L., & Critchlow, B. (1981). Dissonance and alcohol: Drinking your troubles away. *Journal of Personality and Social Psychology, 41*, 831–846.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Sullivan, C. J., & Buckner, C. E. (2005). The influence of perceived role models on college students' purchasing intention and product-related behavior. *Psi Chi Journal of Undergraduate Research, 10*, 66–71.
- Tavris, C. (1992). *The mismeasure of woman*. New York: Touchstone.
- Templer, D. I. (1970). The construction and validation of a death anxiety scale. *Journal of General Psychology, 82*, 165–177.
- Thomas, J., Rewey, K. L., & Davis, S. F. (2002). Professional development benefits of publishing in the *Psi Chi Journal of Undergraduate Research*. *Eye on Psi Chi, 6*, 30–35.
- Torres, A., Zenner, C., Benson, D., Harris, S., & Koberlein, T. (2007). Relationships between self-esteem and factors known to affect college attendance. *Psi Chi Journal of Undergraduate Research, 12*, 204–207.
- Traffanstedt, B. (1998). Weight reduction using behavior modification: Modifying sedentary behavior. *Journal of Psychological Inquiry, 3*, 19–23.
- Tuleya, L. G. (Ed.). (2007). *Thesaurus of psychological index terms* (11th ed.). Washington, DC: American Psychological Association.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–604.
- Vaughn, E. J. (2002, April). *Matching hypothesis and physical attractiveness*. Paper presented at the Arkansas Symposium for Psychology Students, Arkadelphia.
- Walker, K. A., Arruda, J. E., & Greenier, K. D. (1999). Assessment of perceptual biases extracted from the Visual Analogue Scale. *Psi Chi Journal of Undergraduate Research, 4*, 57–63.
- Walter, K. D., Brownlow, S., Ervin, S. L., & Williamson, N. (1998). Something in the way she moves: The influence of shoe-altered gait on motion and trait impressions of women. *Psi Chi Journal of Undergraduate Research, 3*, 163–169.
- Wann, D. L., & Dolan, T. J. (1994). Spectators' evaluations of rival and fellow fans. *The Psychological Record, 44*, 351–358.
- Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist, 44*, 98–104.
- Wells, R. (2001). Susceptibility to illness: The impact of cardiovascular reactivity and the reducer-augmenter construct. *Journal of Psychological Inquiry, 6*, 12–16.
- Welterlin, A. (2004). Social initiation in children with autism: A peer playgroup intervention. *Psi Chi Journal of Undergraduate Research, 9*, 97–104.
- Yang, K. S. (1982). Causal attributions of academic success and failure and their affective consequences. *Chinese Journal of Psychology, 24*, 65–83.
- Zajonc, R. B. (1965). Social facilitation. *Science, 149*, 269–274.
- Zapatel, J., & Garcia-Lopez, S. (2004). Perceptions of lyrics based on music category and perceiver's race. *Psi Chi Journal of Undergraduate Research, 9*, 68–74.
- Zenner, C., & Pritchard, M. (2007). What college students know about breast cancer and eating disorders. *Psi Chi Journal of Undergraduate Research, 12*, 131–134.

APPENDIX A

Statistical Tables

TABLE A-1

The *t* Distribution*

<i>df</i>	α LEVELS FOR TWO-TAILED TEST					
	.20	.10	.05	.02	.01	.001
<i>df</i>	α LEVELS FOR ONE-TAILED TEST					
	.10	.05	.025	.01	.005	.0005
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

*To be significant, the *t* obtained from the data must be equal to or larger than the value shown in the table.

Source: Table A-1 is taken from Table III of Fisher and Yates, *Statistical Tables for Biological, Agricultural and Medical Research*. Published by Longman Group UK Ltd., 1974. We are grateful to the Longman Group UK Ltd., on behalf of the literary Executor of the late Sir Ronald A. Fisher, F.R.S., and Dr. Frank Yates, F.R.S., for permission to reproduce Table III from Statistical Tables for *Biological, Agricultural and Medical Research* 61E (1974).

TABLE A-2 The *F* Distribution***.05 (ROMAN) AND .01 (BOLDFACE) α LEVELS FOR THE DISTRIBUTION OF *F***

DEGREES OF FREEDOM (FOR THE NUMERATOR)																							
1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1 4.052	4.999	5.403	5.625	5.764	5.859	5.981	6.022	6.056	6.082	6.106	6.142	6.169	6.208	6.234	6.258	6.302	6.334	6.352	6.361	6.366		1	
2 18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.38	19.37	19.38	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50	
3 10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.533	
4 7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	
5 6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	
6 5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	
7 5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	
8 5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	
9 5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.71	
10 10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11	5.00	4.92	4.80	4.73	4.64	4.56	4.51	4.45	4.41	4.36	4.33	4.31
11 4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	
12 9.65	7.20	6.22	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40	4.29	4.21	4.10	4.02	3.94	3.86	3.80	3.74	3.70	3.66	3.62	3.60	
13 4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.21	
14 4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	
14 8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.70	3.62	3.51	3.43	3.34	3.26	3.21	3.14	3.11	3.06	3.02	3.00

Degrees of freedom (for the denominator)

TABLE A-2 (Continued)

DEGREES OF FREEDOM (FOR THE NUMERATOR)

	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.12	2.10	2.08	2.07	15	
	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.56	3.48	3.36	3.29	3.20	3.12	3.07	3.00	2.97	2.92	2.89	2.87	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01	16
	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55	3.45	3.37	3.25	3.18	3.10	3.01	2.96	2.89	2.86	2.80	2.77	2.75	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96	17
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45	3.35	3.27	3.16	3.08	3.00	2.92	2.86	2.79	2.76	2.70	2.67	2.65	
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.89	1.95	1.93	1.92	18
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37	3.27	3.19	3.07	3.00	2.91	2.83	2.78	2.71	2.68	2.62	2.59	2.57	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88	19
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.19	3.12	3.00	2.92	2.84	2.76	2.70	2.63	2.60	2.54	2.51	2.49	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84	20
	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23	3.13	3.05	2.94	2.86	2.77	2.69	2.63	2.56	2.53	2.47	2.44	2.42	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81	21
	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17	3.07	2.99	2.88	2.80	2.72	2.63	2.58	2.51	2.47	2.42	2.38	2.36	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78	22
	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.37	2.33	2.31	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.89	1.84	1.82	1.79	1.77	1.76	23
	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.97	2.89	2.78	2.70	2.62	2.53	2.48	2.41	2.37	2.32	2.28	2.26	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73	24
	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36	2.33	2.27	2.23	2.21	
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71	25
	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.32	2.29	2.23	2.19	2.17	
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69	26
	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96	2.86	2.77	2.66	2.58	2.50	2.41	2.36	2.28	2.25	2.19	2.15	2.13	
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67	27
	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.14	3.06	2.98	2.74	2.63	2.55	2.47	2.38	2.33	2.25	2.21	2.16	2.12	2.10			
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65	28
	7.64	5.45	4.57	4.07	3.46	3.53	3.36	3.23	3.11	3.03	2.95	2.90	2.80	2.71	2.60	2.52	2.44	2.35	2.30	2.22	2.18	2.13	2.09	2.06	
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64	29
	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.08	3.00	2.92	2.87	2.77	2.68	2.57	2.49	2.41	2.32	2.27	2.19	2.15	2.10	2.06	2.03	

Degrees of freedom (for the denominator)

(Continued)

TABLE A-2 (Continued)

DEGREES OF FREEDOM (FOR THE NUMERATOR)

	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62	30
	756	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.90	2.84	2.74	2.66	2.55	2.47	2.38	2.29	2.24	2.16	2.13	2.07	2.03	2.01	
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59	32
	750	5.34	4.46	3.97	3.66	3.42	3.25	3.12	3.01	2.94	2.86	2.80	2.70	2.62	2.51	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96	
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57	34
	744	5.29	4.42	3.93	3.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76	2.66	2.58	2.47	2.38	2.30	2.21	2.15	2.08	2.04	1.98	1.94	1.91	
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55	36
	739	5.25	4.38	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72	2.62	2.54	2.43	2.35	2.26	2.17	2.12	2.04	2.00	1.94	1.90	1.87	
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53	38
	735	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69	2.59	2.51	2.40	2.32	2.22	2.14	2.08	2.00	1.97	1.90	1.86	1.84	
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	40	
	731	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66	2.56	2.49	2.37	2.29	2.20	2.11	2.05	1.97	1.94	1.88	1.84	1.81	
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49	42
	727	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64	2.54	2.46	2.35	2.26	2.17	2.08	2.02	1.94	1.91	1.85	1.80	1.78	
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48	44
	724	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62	2.52	2.44	2.32	2.24	2.15	2.06	2.00	1.92	1.88	1.82	1.78	1.75	
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46	46
	721	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60	2.50	2.42	2.30	2.22	2.13	2.04	1.98	1.90	1.86	1.80	1.76	1.72	
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45	48
	719	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58	2.48	2.40	2.32	2.20	2.11	2.02	1.96	1.88	1.84	1.78	1.73	1.70	
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.70	1.65	1.63	1.60	1.55	1.52	1.48	1.46	50
	717	5.06	4.20	3.72	3.41	3.18	3.02	2.88	2.78	2.70	2.62	2.56	2.46	2.39	2.26	2.18	2.10	2.00	1.94	1.86	1.82	1.76	1.71	1.68	
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41	55
	712	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53	2.43	2.35	2.23	2.15	2.06	1.96	1.90	1.82	1.78	1.71	1.66	1.64	
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39	60
	708	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.40	2.32	2.20	2.12	2.03	1.93	1.87	1.79	1.74	1.68	1.63	1.60	
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37	65
	704	4.95	4.10	3.62	3.31	3.09	2.93	2.79	2.70	2.61	2.54	2.47	2.37	2.30	2.18	2.09	2.00	1.90	1.84	1.76	1.71	1.64	1.60	1.56	
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35	70
	701	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45	2.35	2.28	2.15	2.07	1.98	1.88	1.82	1.74	1.69	1.62	1.56	1.53	

Degrees of freedom (for the denominator)

TABLE A-2 (Continued)

DEGREES OF FREEDOM (FOR THE NUMERATOR)

	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32	80
	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.41	2.32	2.24	2.11	2.03	1.94	1.84	1.78	1.70	1.65	1.57	1.52	1.49	
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28	100
	6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36	2.26	2.19	2.06	1.98	1.89	1.79	1.73	1.64	1.59	1.51	1.46	1.43	
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25	125
	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.65	2.56	2.47	2.40	2.33	2.23	2.15	2.03	1.94	1.85	1.75	1.68	1.59	1.54	1.46	1.40	1.37	
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22	150
	6.81	4.75	3.91	3.44	3.14	2.92	2.76	2.62	2.53	2.44	2.37	2.30	2.20	2.12	2.00	1.91	1.83	1.72	1.66	1.56	1.51	1.43	1.37	1.33	
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19	200
	6.76	4.71	3.88	3.41	3.11	2.90	2.73	2.60	2.50	2.41	2.34	2.28	2.17	2.09	1.97	1.88	1.79	1.69	1.62	1.53	1.48	1.39	1.33	1.28	
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13	400
	6.70	4.66	3.83	3.36	3.06	2.85	2.69	2.55	2.46	2.37	2.29	2.23	2.12	2.04	1.92	1.84	1.74	1.64	1.57	1.47	1.42	1.32	1.24	1.19	
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08	1000
	6.66	4.62	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.26	2.20	2.09	2.01	1.89	1.81	1.71	1.61	1.54	1.44	1.38	1.28	1.19	1.11	
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00	∞
	6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.24	2.18	2.07	1.99	1.87	1.79	1.69	1.59	1.52	1.41	1.36	1.25	1.15	1.00	

*To be significant, the F obtained from the data must be equal to or larger than the value shown in the table.

Source: Statistical Methods (6th ed.) by G. W. Snedecor and W. G. Cochran, 1967, Ames, Iowa: Iowa State University Press. Used by permission of the Iowa State University Press.

TABLE A-3

Critical values for Pearson product-moment correlation coefficients, r . To be significant, the obtained r must be equal to or larger than the table value.

df	α LEVELS (TWO-TAILED TEST)				
	.1	.05	.02	.01	.001
$(df = N - 2)$	α LEVELS (ONE-TAILED TEST)				
	.05	.025	.01	.005	.0005
1	.98769	.99692	.999507	.999877	.9999988
2	.90000	.95000	.98000	.990000	.99900
3	.8054	.8783	.93433	.95873	.99116
4	.7293	.8114	.8822	.91720	.97406
5	.6694	.7545	.8329	.8745	.95074
6	.6215	.7067	.7887	.8343	.92493
7	.5822	.6664	.7498	.7977	.8982
8	.5494	.6319	.7155	.7646	.8721
9	.5214	.6021	.6851	.7348	.8371
10	.4973	.5760	.6581	.7079	.8233
11	.4762	.5529	.6339	.6835	.8010
12	.4575	.5324	.6120	.6614	.7800
13	.4409	.5139	.5923	.6411	.7603
14	.4259	.4973	.5742	.6226	.7420
15	.4124	.4821	.5577	.6055	.7246
16	.4000	.4683	.5425	.5897	.7084
17	.3887	.4555	.5285	.5751	.6932
18	.3783	.4438	.5155	.5614	.6787
19	.3687	.4329	.5034	.5487	.6652
20	.3598	.4227	.4921	.5368	.6524
25	.3233	.3809	.4451	.4869	.5974
30	.2960	.3494	.4093	.4487	.5541
35	.2746	.3246	.3810	.4182	.5189
40	.2573	.3044	.3578	.3932	.4896
45	.2428	.2875	.3384	.3721	.4648
50	.2306	.2732	.3218	.3541	.4433
60	.2108	.2500	.2948	.3248	.4078
70	.1954	.2319	.2737	.3017	.3799
80	.1829	.2172	.2565	.2830	.3568
90	.1726	.2050	.2422	.2673	.3375
100	.1638	.1946	.2301	.2540	.3211

Source: Table VII of Fisher and Yates (1963), *Statistical Tables for Biological, Agricultural and Medical Research*, published by Longman Group, Ltd., London.

To help refresh your memory and understanding of the major statistical tests mentioned in the text, we have included complete formulae in this appendix. In all but one instance we present the raw-score formulae. (The one exception is the calculation of variance, where we present the deviation-score formula.)

In all instances, we have attempted to present the simplest formulae. If they look confusing or unfamiliar, consult your instructor; there are numerous ways to write these formulae. The meaning of statistical symbols and notations are presented where they might be confusing or unfamiliar. With these formulae you should be able to work any of the examples from the text by hand.

I. Variance and Standard Deviation

Sample

$$\text{Variance} = \frac{\sum(X - \bar{X})^2}{N - 1}$$

$$\text{Standard Deviation} = \sqrt{\text{Variance}}$$

Population

$$\text{Variance} = \frac{\sum(X - \bar{X})^2}{N}$$

$$\text{Standard Deviation} = \sqrt{\text{Variance}}$$

II. Pearson Product-Moment Correlation Coefficient

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

where N = the number of *pairs* of scores

III. *t* Test for Independent Samples

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{\sum X_1^2 - (\sum X_1)^2}{N_1} \right] + \left[\frac{\sum X_2^2 - (\sum X_2)^2}{N_2} \right] \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

IV. *t* Test for Related Samples

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{N}}{N-1} \left(\frac{1}{N} \right)}}$$

where D = difference between measurement 1 and measurement 2

N = the number of *pairs* of scores

V. One-Way Analysis of Variance

Between Groups Sum of Squares

$$SS_{BG} = \left[\frac{(\sum X_1)^2}{N_1} + \frac{(\sum X_2)^2}{N_2} + \dots + \frac{(\sum X_j)^2}{N_j} \right] - \frac{(\sum X_{Total})^2}{N_{Total}}$$

Within Groups Sum of Squares

$$SS_{WG} = \left[\sum X_1^2 - \frac{(\sum X_1)^2}{N_1} \right] + \left[\sum X_2^2 - \frac{(\sum X_2)^2}{N_2} \right] + \dots + \left[\sum X_j^2 - \frac{(\sum X_j)^2}{N_j} \right]$$

Total Sum of Squares

$$SS_{Total} = \sum X_{Total}^2 - \frac{(\sum X_{Total})^2}{N_{Total}}$$

Factorial Design with Three Independent Variables

Three-Way ANOVA for Independent Samples. As we mentioned in Chapter 12, the possibilities for factorial designs are almost limitless when you consider that you can vary the number of IVs, vary the number of levels for each IV, and vary how participants are assigned to groups for each IV. Thus, there is no way that we could cover every possible type of factorial design in Chapter 12. We concentrated on the simplest factorial designs (2×2) in order to provide a basic introduction. This Appendix gives an example of a factorial design involving three IVs. A three-way ANOVA for independent samples involves three IVs (thus the term *three-way*) that all use independent groups of participants.

To implement this design, we will make one final alteration to our first experiment involving clothing and customer sex in Chapter 12 (using the factorial between-groups design). Imagine that you are discussing your results from that experiment with a friend (remember that we found a significant interaction between clothing and customer sex such that men who were dressed in sloppy clothes received help from salesclerks most slowly). Your friend wonders aloud whether that result would hold true for all types of stores. (This is how experimental ideas develop sometimes!) That musing piques your curiosity, and you set out to test this new question. Trying to keep your design simple, you decide to contrast two different types of stores (a large discount store vs. an exclusive clothing store). Thus, your experimental design would now be a $2 \times 2 \times 2$ (clothing by customer sex by store type) completely randomized design (all groups composed of different participants). The general design for this experiment is depicted in Figure 12-11 on page 279. The response times of the clerks to the customers in the various treatment combinations appear in Table C-1. The means for the variables and their combinations also appear in Table C-1.

Computer Results. The source table for this analysis appears in Table C-2. As you can see, the table contains a wealth of information, neatly arranged by main effects, two-way interactions, and the three-way interaction. Remember we have talked many times about the notion that interactions qualify main effects. When we look at the three-way interaction, we find that it is significant ($p = .001$). If it had *not* been significant, then we would have examined and graphed the two-way interactions. (Although two of them—CLOTHES by STORE and CSEX by STORE—are significant, we do not interpret them because the three-way interaction supersedes them.) If the three-way interaction was not significant and none of the two-way interactions were significant, then we could have interpreted the main effects in a straightforward manner. Although CLOTHES and STORE show significance, we should not attempt to interpret them because the three-way interaction is significant. To interpret the CLOTHES by CSEX by STORE interaction, we must graph it (see Figure C-1).

Translating Statistics Into Words. We found a significant three-way interaction. All other main effects and interactions are qualified because of this interaction. In Figure C-1, we see a classic interaction pattern as the lines on the graph cross.

From examining Figure C-1 carefully, it appears that salesclerks' response time to customers depended on the clothing and sex of the customer *and* the store in which the clerks worked. One line shows an obvious difference from the other three—the line for male customers in

TABLE C-1 Clerk Response Time for Clothing by Customer Sex by Type of Store

		EXCLUSIVE CLOTHING STORE		LARGE DISCOUNT STORE		
		Customer Sex		Customer Sex		
		Women	Men	Women	Men	
CLOTHING	Casual	46 39 50 <i>M = 48.17</i>	38 50 38 <i>M = 45.67</i>	66 59 70 <i>M = 61.50</i>	58 70 58 <i>M = 65.67</i>	Casual <i>M = 55.25</i>
	Casual	52 48 54	44 49 55	72 48 54	64 69 75	
	Sloppy	37 47 44 <i>M = 51.50</i>	47 69 69 <i>M = 68.67</i>	57 67 64 <i>M = 71.50</i>	61 52 54 <i>M = 58.50</i>	Sloppy <i>M = 62.54</i>
	Sloppy	62 49 70	74 77 76	82 69 90	59 67 58	
		Women <i>M = 58.17</i>	Men <i>M = 59.63</i>	Exclusive Clothing Store <i>M = 53.50</i>	Large Discount Store <i>M = 64.29</i>	
				Total Population <i>M = 58.90</i>		

sloppy clothes. Notice that all three of the other groups (casually dressed men, casually dressed women, and sloppily dressed women) received slower service in the large discount store than in the exclusive clothing store. For sloppily dressed men, however, the pattern was reversed—they received faster attention in the discount store. We are not sure why these results occurred; we can only speculate. The pattern for the three similar groups probably depicts the experience most of us have had: It is far easier to find a salesclerk to help you in a “fancy” store than in a large discount store. On the other hand, perhaps clerks in the exclusive clothing store want nothing to do with a sloppily dressed man, so they ignore him for longer periods of time. In the large discount store, perhaps such a man gets help a little more quickly because

TABLE C-2 Computer Output for Three-Way ANOVA for Independent Samples

SOURCE TABLE					
SOURCE	SS	DF	MS	F	P
CLOTHES	638.02	1	638.02	7.69	.008
CSEX	25.52	1	25.52	.31	.582
STORE	139752	1	139752	16.85	.001
CLO × CSEX	4.69	1	4.69	.06	.813
CLO × STORE	414.19	1	414.19	4.99	.031
CSEX × STORE	414.19	1	414.19	4.99	.031
CLO × CSEX × STORE	101752	1	101752	12.27	.001
Residual	3316.83	40	82.92		
TOTAL	7228.48	47			

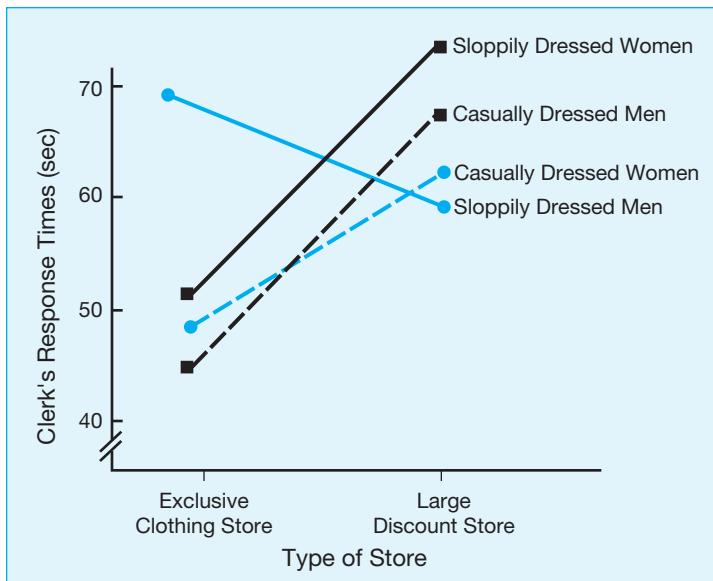


FIGURE C-1

the clerks are suspicious that he might be a shoplifter. Again, this is only conjecture—the type of hypothesizing we would save for our discussion section.

One last note about this design: We need to communicate our findings in APA style. You may shudder at the thought of writing such a large number of results in an APA-format paragraph. Fortunately, with larger designs such as this, researchers often take a shortcut when writing their results. Here's an example of a short way to communicate these findings.

The salesclerks' response times were analyzed with a three-way randomized ANOVA, using clothing, customer sex, and type of store as independent variables. The interaction of clothing by customer sex by store type was significant, $F(1, 40) = 12.27, p = .001 \eta^2 = .23$, and is graphed in Figure 1 [see Figure C-1]. All main effects and two-way interactions were qualified by the three-way interaction.

As you can see, we spent the bulk of our commentary on our significant effect and summarized the remaining six effects (three main effects and three two-way interactions) in a single sentence. This type of coverage is not only economical but also focuses on the important finding—the one that was significant and rendered the other significant effects moot. Remember in our discussion in Chapter 12 we pointed out how an interaction qualifies the main effects. Although our source table for this three-way ANOVA shows significant effects for both CLOTHES and STORE, neither makes sense by itself. We can't say that clerks react more quickly to customers dressed in casual clothes or that clerks in exclusive clothing stores respond to customers more quickly; the three-way interaction makes us realize that an explanation based solely on clothing or store type is too simplistic to describe the results that we found.

You may also remember that we said in Chapter 12 that factorial designs rarely use more than three or four IVs at a time. You have seen a three-way interaction in this Appendix; imagine adding a fourth variable to the interaction! Still, factorial experiments using two or three variables are quite popular with researchers because of the wealth of information they can provide. We urge you to consider such a design as you plan your research.

Chapter 1

CHECK YOUR PROGRESS PAGE 13

- 1.** The steps involved in the research process include the following:

Finding a Problem. You detect a gap in the knowledge base.

Literature Review. You consult previous reports to determine what has been found in the research area of interest.

Theoretical Considerations. The literature review highlights theories that point to relevant research projects.

Hypothesis. The literature review also highlights hypotheses (statements of the relationship between variables in more restricted domains of the research area). Such hypotheses will assist in the development of the experimental hypothesis—the predicted outcome of your research project.

Research Design. You develop the general plan for conducting the research project.

Conducting the Research. You conduct the research project according to the experimental design.

Analysis of Research Findings. Based on the results of your data analysis, you will decide the importance (*significance*) of your research findings.

Decisions in Terms of Past Research and Theory. The analysis of research findings guides your decisions concerning the relationship of your present research project to past research and theoretical considerations.

Preparation of the Research Report. You prepare a research report describing the rationale, conduct, and results of the experiment according to accepted American Psychological Association (APA) format.

Sharing Your Results: Presentation and Publication. You share your research report with colleagues at a professional society meeting and/or by publication in a professional journal.

Finding a New Problem. Your research results unearth another gap in our knowledge base and the research process begins again.

2. d

3. d

4. d

5. b

6. d

- 7** The reasons for taking a research methods or experimental psychology course include that it will
- assist you in other psychology courses
 - enable you to conduct a research project after graduation
 - assist you in getting into graduate school
 - help you to become a knowledgeable consumer of psychological research

Chapter 2

CHECK YOUR PROGRESS PAGE 28

- 1.** 1-C; 2-A; 3-D; 4-B; 5-E
- 2.** testable; likelihood of success
- 3.** Nonsystematic sources of research ideas include those instances or occurrences that make us believe that the research idea was unplanned. These types of research ideas are shown through "inspiration," "serendipity," and "everyday occurrences." Systematic sources of research ideas are carefully organized and logically thought out. This source of research ideas may be found in "past research," "theory," and "classroom lectures."
- 4.** c
- 5.** d
- 6.** a
- 7.** c

CHECK YOUR PROGRESS PAGE 47

- 1.** During World War II, atrocities took place in which prisoners of war were unwillingly subjected to experiments using drugs, viruses, and toxic agents. The Tuskegee syphilis study examined the course of untreated syphilis in a group of men who did not realize they were being studied. Doctors in the Willowbrook experiment purposely infected patients with hepatitis and did not give them treatment, in order to study the development of the disease. Milgram's research on obedience studied the reaction to demands from an authority, in participants who were being deceived.
- 2.** The Nuremberg Code stressed the need to
 - obtain participants' consent to participate in research
 - inform participants about the nature of the research project
 - avoid risks where possible
 - protect participants against risks
 - conduct research using qualified personnel
- 3.** Deception may be needed if the results of a research project would be biased or contaminated by the participants knowing the nature of the experiment. A more general statement on the informed consent document in conjunction with a clear indication

that the participants may terminate the project at any time might be used. Complete debriefing should follow the experiment.

4. a

5. d

6. The guidelines for the ethical use of animals in research involve attention to each of the following:

I. Justification of Research. The research should have a clear scientific purpose.

II. Personnel. Only trained personnel who are familiar with the animal care guidelines should be involved with the research. All procedures must conform to appropriate federal guidelines.

III. Care and Housing of Animals. Animal housing areas must comply with current regulations.

IV. Acquisition of Animals. If animals are not bred in the laboratory, they must be acquired in a lawful, humane manner.

V. Experimental Procedures. "Humane consideration for the well-being of the animal should be incorporated into the design and conduct of all procedures involving animals, while keeping in mind the primary goal of experimental procedures—the acquisition of sound, replicable data."

VI. Field Research. Field research must be approved by the appropriate review board. Investigators should take special precautions to disturb their research population(s) and the environment as little as possible.

VII. Educational Use of Animals. The educational use of animals must be approved by the appropriate review board. Instruction in the ethics of animal research is encouraged.

7. The IRB is the Institutional Review Board. The typical IRB at a college or university is composed of faculty members from a variety of disciplines, members from the community, and a veterinarian if animal proposals are considered. The IRB examines proposed procedures, questionnaires, the informed consent document, debriefing plans, the use of pain in animals, and proposed procedures for animal disposal.

8. The experimenter is responsible for the ethical conduct of the research project and the ethical presentation of the research results.

9. Plagiarism refers to the use of someone else's work without giving credit to the original author. Fabrication of data involves the creation of fictitious research data. Pressures for job security (tenure), salary increases, and ego involvement in the research are factors that might lead to such unethical behaviors.

Chapter 3

CHECK YOUR PROGRESS PAGE 57

1. Qualitative research seeks to understand social or human issues and problems; seeks to develop a holistic pictures of these problems; creates its report with words, not statistics; uses the views of informants as the basis for its report; and, is conducted in a natural setting.

- 2.** b
- 3.** b
- 4.** d
- 5.** d
- 6.** a
- 7.** a
- 8.** d

Chapter 4

CHECK YOUR PROGRESS PAGE 69

- 1.** The reactance or reactivity effect refers to the biasing or influencing of the participants' scores or responses because they know they are being observed. Archival research avoids the reactance effect because the researcher does not observe the participants and the data of interest are recorded before they are used in the research project.
- 2.** d
- 3.** a
- 4.** b
- 5.** Situation sampling and time sampling are used to provide greater generality for the research project.
- 6.** The extent to which two observers agree is called interobserver reliability. It is calculated by dividing the number of times the observers agree by the number of opportunities for agreement and multiplying by 100. These calculations result in the percent agreement.
- 7.** c

CHECK YOUR PROGRESS PAGE 84

- 1.** 1-D; 2-F; 3-G; 4-C; 5-A; 6-E; 7-B
- 2.** The steps involved in developing a good survey include (a) determining how the information you seek is to be obtained and what type of instrument will you use, (b) determining the nature of the questions that will be used, (c) writing the items for your survey, (d) pilot testing your survey or questionnaire, (e) considering what demographic data are desired, and (f) specifying the procedures that will be followed in administering the survey or questionnaire.
- 3.** d
- 4.** The low return rate of mail surveys can be increased by (a) including a letter that clearly summarizes the nature and importance with the initial mailing (a self-addressed, pre-paid return envelope should be included with the initial mailing) and (b) sending additional mailings at two- to three-week intervals.
- 5.** The use of personal interviews is declining because (a) the expense involved in conducting them has increased greatly, (b) the fact that an individual administers the surveys

increases the possibility for bias, and (c) the appeal of going from door to door is decreasing because of unavailability of respondents and high crime rates in large metropolitan areas.

6. Achievement tests are used to assess an individual's level of mastery or competence. Aptitude tests are used to assess an individual's skills or abilities.
7. d
8. population; sample
9. b
10. Random sampling occurs when every member of the population has an equal chance of being selected. Random sampling with replacement occurs when the chosen items are returned to the population and can be selected on future occasions. When random sampling without replacement is used, the chosen item is not returned to the population.
11. Stratified random sampling involves randomly selecting participants from a single layer or strata in the population (e.g., a limited age range). When stratified random sampling is used, group homogeneity is increased. The more homogeneous the sample, the fewer chances there are for nuisance variables to operate. The fewer chances there are for nuisance variables to operate, the smaller the within-group variability.
12. The single-strata research approach attempts to secure research data from a single, specified segment of the population of interest. The cross-sectional research approach involves the comparison of two or more groups of participants during the same rather limited time span. The longitudinal research approach involves gathering information from a group of participants over an extended period of time.

Chapter 5

CHECK YOUR PROGRESS PAGE 98

1. The components of the scientific method are **objectivity** (unbiased collection of data), **confirmation of findings** (the ability to repeat or replicate research projects), **self-correction** (replication of research removes erroneous findings), and **control** (accounting for the effects of unwanted factors and direct manipulation of the independent variable).
2. Self-correction means that the replication of scientific research results in the elimination of errors and erroneous findings.
3. b
4. The experimenter attempts to determine the relation between the independent variable (cause) and subsequent changes in the dependent variable (effect).
5. b
6. d
7. Synthetic statements are used in the experimental hypothesis because they can be either true or false.
8. General implication form presents the experimental hypothesis as an "if . . . then" statement where the "if" portion of the statement refers to the IV manipulation and the "then" portion of the statement refers to the predicted changes in the DV.

9. d**10.** inductive, deductive**11.** b

Chapter 6

CHECK YOUR PROGRESS PAGE 112

1. variable**2.** 1-E; 2-C; 3-A; 4-F; 5-B; 6-D**3.** c**4.** c**5.** You should record more than one DV if the additional DV(s) add meaningful information.**6.** c

CHECK YOUR PROGRESS PAGE 124

1. 1-C; 2-A; 3-B; 4-E; 5-D**2.** a**3.** a**4.** When within-subject counterbalancing is used, each participant experiences more than one sequence of IV presentations. When within-group counterbalancing is used, each participant experiences a different sequence of IV presentations.**5.** $n!$ refers to factoring or breaking a number into its component parts and then multiplying these component parts. $n!$ can be used to determine the number of sequences required for complete counterbalancing: $4! = 4 \times 3 \times 2 \times 1 = 24$.**6.** b**7.** The incomplete counterbalancing procedure refers to the use of some, but not all, of the possible sequences of treatment administration.

Chapter 7

CHECK YOUR PROGRESS PAGE 132

1. Once a precedent or established pattern for using a particular type of research participant is begun, it is likely that that type of participant will be used in experiments in the research area in question.**2.** White rats and college students are the favorite participants in psychological research because of precedent and availability. An established pattern of research with these two populations has been established, and they are easy to obtain.**3.** b

4. c

5. a

6. The experimenter must be careful not to become a slave to elaborate pieces of equipment. If this situation occurs, then it is likely that the equipment may begin dictating the type of research that is conducted and/or the type of DV that is recorded.

CHECK YOUR PROGRESS PAGE 140

1. The experimenter's physiological characteristics, psychological characteristics, and personal expectancies for the outcome of the experiment can operate as extraneous variables and influence the responses of the participants.

2. 1-C; 2-A; 3-B; 4-E; 5-D

3. a

4. Automated equipment and instruments to present instructions and record data are frequently used to control for experimenter expectancies because they help minimize experimenter contact with the participants. By minimizing contact with the participants, the experimenter is less likely to influence the outcome of the experiment.

5. b

6. c

CHECK YOUR PROGRESS PAGE 146

1. a

2. c

3. The goal of cross-cultural psychology is incompatible with ethnocentrism because ethnocentrism views other cultures as an extension of that culture. Hence, according to an ethnocentric view, there is no need for cross-cultural research.

4. Culture can influence the choice of the research problem, the nature of the experimental hypothesis, selection of the IV(s), selection of the DV(s), selection of participants, sampling procedures, and the type of questionnaire that is used.

5. b

Chapter 8

CHECK YOUR PROGRESS PAGE 156

1. a

2. b

3. It is important to evaluate your experiment for internal validity because you cannot place any confidence in your results if your experiment does not have internal validity. Cause-and-effect statements cannot be made without internal validity.

4. 1-B; 2-C; 3-D; 4-A

5. 1-D; 2-A; 3-B; 4-C

- 6.** In this experiment the internal validity threat of selection is likely. Because the selected participants are senior citizens, it is possible that they will have less formal education than college students of today. Obtaining a college education was not as common for today's senior citizens when they were young as it is for today's youth.

7. b

CHECK YOUR PROGRESS PAGE 170

- 1.** External validity is the ability to take your experimental findings and apply them beyond the experimental participants to the larger population. It is important to psychology so that we can develop general findings that apply to large groups of organisms.
- 2. b**
- 3.** Population generalization involves applying results from participants to the larger group. Environmental generalization involves applying results to different settings from those in the original experiment. Temporal generalization involves applying experimental results to different times from those in the original experiment.
- 4.** Using different types of participants helps increase the external validity of our findings.
- 5.** Cross-cultural psychology involves testing psychological principles in different cultures to determine the generality of those principles. It is relevant to this chapter because it deals with external validity.
- 6.** 1-C; 2-B; 3-D; 4-A
- 7.** Mook thinks that external validity is necessary only when we are attempting to predict behavior in the real world. Much research in psychology is not aimed at such prediction, and Mook believes that external validity is not necessary in such situations.
- 8.** It is almost impossible to achieve external validity in a single experiment because of the large number of threats to external validity. Usually an experiment can answer only one such threat at a time.

Chapter 9

CHECK YOUR PROGRESS PAGE 187

- 1.** 1-G; 2-E; 3-A; 4-C; 5-B; 6-F; 7-D
- 2. c**
- 3.** Because the three measures of central tendency (mean, median, and mode) are identical in a normal distribution, any one of them would serve as a representative measure of central tendency.
- 4. d**
- 5.** The ordinate is the vertical axis; it should be approximately two-thirds the size of the abscissa (horizontal axis). The DV is plotted on the ordinate, whereas the IV is plotted on the abscissa.
- 6.** Because it takes into account only the highest and lowest scores and disregards the distribution of scores in between, the range does not convey much information.
- 7. a**

- 8.** Because the percentage of scores occurring between the mean and any standard deviation (e.g., 11, 12, -1, -2) away from the mean is constant, standard deviation scores from one distribution can be compared with standard deviation scores from other distributions.

CHECK YOUR PROGRESS PAGE 201

- 1.** 1-F; 2-H; 3-I; 4-B; 5-D; 6-C; 7-G; 8-E; 9-A
- 2.** A positive correlation indicates that as one variable increases, the other variable under consideration also increases, whereas a perfect positive correlation indicates that for every unit of increase in one variable there is a corresponding increase of one unit in the other variable.
- 3.** A zero correlation indicates that changes in one variable are not systematically related to changes in the other variable.
- 4.** When two independent groups are being compared, the *t* test compares the difference between the means of the two groups to the amount of variability (error) that exists within the two groups. Because the groups are assumed to be equivalent at the start of the experiment, larger *t* values indicate greater influence of the IV.
- 5.** “Level of significance” refers to the point at which the experimenter feels that a result occurs rarely by chance. If an experimental result occurs rarely by chance, we conclude that it is significant. Although the experimenter arbitrarily sets the level of significance, tradition has established the .05 level as an accepted level of significance.
- 6.** Because researchers don’t always know how a research project is going to turn out, it is safer to state a nondirectional hypothesis and use a two-tail test. If a directional hypothesis is stated and the results turn out differently, then the researcher is forced to reject the experimental hypothesis even though differences may exist between the groups.
- 7.** a
- 8.** a
- 9.** c

Chapter 10

CHECK YOUR PROGRESS PAGE 214

- 1.** We cannot conduct a valid experiment with only one group because we have to have a second group for comparison purposes. We cannot tell whether our IV has any effect if we cannot compare the experimental group to a control group.
- 2.** levels
- 3.** Independent groups consist of participants who are totally unrelated to each other. Correlated groups are composed of pairs of participants who have some relationship to each other because they (a) have some natural relationship (natural pairs), (b) are matched with each other (matched pairs), or (c) are the same participants (repeated measures). This difference is important to experimental designs because it is one of the three questions we ask to choose an appropriate design for an experiment.
- 4.** 1-D; 2-A; 3-B; 4-C
- 5.** We must be cautious about the use of random assignment when we have small numbers of participants because the groups may end up being unequal before the experiment.

- 6.** If your experiment could use either independent or correlated groups, you would most likely base your decision on the number of participants available to you. We often use correlated groups when we have a small number of participants so that we can be more confident about the equality of our groups. If you have many potential participants, independent groups (through random assignment) will typically be equal.

7. b

CHECK YOUR PROGRESS PAGE 220

- 1.** It is important for our groups to be equal before the experiment so that any difference we detect in our DV can be attributed to the IV.
- 2.** between-groups variability; error variability
- 3. b**
- 4.** Correlated groups have the advantages of ensuring equality of the participants before the experiment and of reducing error variability. Independent groups are advantageous in that they are simpler than correlated groups and can be used in experimental situations that preclude correlated groups.
- 5.** Many different answers are possible to the question about comparing differing amounts of an IV. Some representative examples include amount of study time or amount of reinforcement in a learning experiment, amount of pay or bonus on job performance, and length of therapy to treat a particular problem. As long as you chose a single IV and varied its amount (rather than presence vs. absence), your answer should be correct.
- 6.** Again, many different answers are possible. Some possible answers include different types of life experiences, different majors, different musical preferences, and different hometowns. As long as you chose two levels of an IV that cannot be manipulated, your answer is correct.

CHECK YOUR PROGRESS PAGE 230

- 1.** To compare the stereotyping of a group of female executives to the stereotyping of a group of male executives, you would use a *t* test for independent samples because men and women represent independent groups.
- 2.** To compare the stereotyping of a group of male executives before and after the ERA, you would use a *t* test for correlated samples because this represents repeated measures.
- 3. a**
- 4.** We usually look for descriptive statistics first on a computer printout because this information helps us understand how the groups performed on the DV.
- 5. homogeneity of variance; heterogeneity of variance**
- 6. words; numbers (statistical information)**
- 7.** Group A, with a mean of 75, scored significantly higher than Group B, which had a mean of 70.
- 8.** Research is a cyclical, ongoing process because each experiment typically raises new questions. A multitude of examples of cyclical research could be given. For example, you might test a new antihyperactivity drug against a control group. If the drug were helpful, then you would need additional research to determine the most effective dosage. In the future you would want to test this drug against new drugs that arrive on the market.

Chapter 11

CHECK YOUR PROGRESS PAGE 243

1. The two-group design is the building block for the multiple-group design because the multiple-group design essentially takes a two-group design and adds more groups to it. Thus, the multiple-group design is similar to changing a two-group design into a three-group design (or larger).
2. one; three
3. A multiple-group design allows you to ask and answer more questions than does a two-group design. You may be able to run only one experiment instead of two or three. Therefore, the multiple-group design is more efficient than the two-group design—you can save time, effort, participants, and so on.
4. There are many correct answers to this question. If you chose an IV that has more than two levels, your answer should be correct. For example, if you wished to test attendance in college students as a function of classification, you would use a multiple-group design with four groups (freshmen, sophomores, juniors, seniors).
5. Matched sets, repeated measures, and natural sets are all considered correlated groups because participants in such groups are related in some way.
6. There is no real limit on the number of groups in a multiple-group design. Practically speaking, it is rare to see more than four or five groups for a particular IV.
7. How many groups are required to adequately test your experimental hypothesis? Will you learn something important if you include more levels of your IV?
8. provide more control than independent-groups designs
9. Practical considerations are more demanding in multiple-correlated-groups designs because there are simply more “equated” participants with which to deal. In repeated measures, participants must take part in more experimental sessions. Matched sets must include three (or more) matched participants. Natural sets require larger sets of participants.
10. To compare personality traits of firstborn, lastborn, and only children, we would use a multiple-independent-groups design (unless we matched participants on some variable). Repeated measures and natural sets are impossible (can you figure out why?). This would represent an ex post facto study because we cannot manipulate an individual’s birth order.

CHECK YOUR PROGRESS PAGE 257

CLASSIFICATION			
Freshman	Sophomore	Junior	Senior

This experiment would require a multiple-independent-groups design with four groups. The necessary statistical test would be a one-way ANOVA for independent groups (students could not be in more than one classification simultaneously).

2.

ACT/SAT ATTEMPTS

1st Attempt

2nd Attempt

3rd Attempt

--	--	--

This question requires a multiple-correlated-groups design with three groups. The proper statistical test would be a one-way ANOVA for correlated groups because each student takes the test three times (repeated measures).

3. Because you have more than two groups, it is necessary to compute post hoc tests to tell which groups are significantly different from the others.
4. between groups; within groups
5. Given the statistical information shown, you can conclude that students perform differently on their three attempts at taking the ACT or SAT (i.e., there is a statistically significant difference).
6. To draw a full and complete conclusion from Question 5, you would need the results from post hoc tests comparing the three means. With this information you could conclude whether students improve over time in taking the entrance exam.
7. It was important to have the information from Chapter 10's continuing research problem so that you could begin a follow-up experiment with the knowledge that customers' clothing affects clerks' response times. The follow-up experiment, then, compared clerks' response times to three different clothing styles.
8. It depends. Do you want to measure the same people's moods across the four seasons (multiple-correlated-groups design) or four different groups of people in the four seasons (multiple independent groups)? Either approach is possible. Can you justify your answer? If you chose multiple correlated groups, your rationale should revolve around control issues. You should measure a large number of participants if you chose multiple independent groups.
9. It depends. Do you want the same people to eat at all four restaurants (multiple correlated groups) or to survey different people at each restaurant (multiple independent groups)? You could run the experiment either way. The rationale for your choice should be similar to that summarized in the answer to Question 8.

Chapter 12

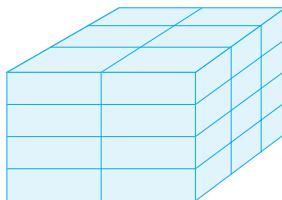
CHECK YOUR PROGRESS PAGE 273

1. The two-group design is related to the factorial design because it forms the underlying basis of a factorial design. For example, a 2×2 factorial design is simply two two-group designs combined (see the shaded section below).

- 2.** There is a practical limit to the number of IVs you can use in an experiment so that you will be able to interpret the results easily. Interactions involving many variables can be quite difficult to understand.
- 3.** d
- 4.** 1-C; 2-A; 3-B
- 5.** two; four
- 6.** Numerous correct answers are possible. Your answer should consist of an experiment with two IVs. One IV should be a between-subjects variable (no relationship between participants); one should be a within-subjects variable (repeated measures, matching, or natural groups). For example, you might assign students to either a group that takes an ACT or SAT preparation course or does not take such a course (between subjects). Each group would take the ACT or SAT twice (repeated measures).

CHECK YOUR PROGRESS PAGE 282

- 1.** Factorial designs are combinations of the designs in Chapters 10 and 11 because a factorial design results from adding two (or more) simple designs (one IV each) together into a single design with two (or more) IVs (e.g., see Figure 12-10, page 274).
- 2.** A $2 \times 4 \times 3$ experimental design consists of three IVs; one has two levels, one has four levels, and one has three levels:



- 3.** Totally between-groups designs use independent groups of participants for each IV. Totally within-groups designs use correlated groups of participants for all IVs. Mixed-groups designs have at least one IV that uses independent groups of participants and one that uses correlated groups. These designs are similar in that they are all factorial designs. They differ, of course, in the way that the experimenter assigns participants to groups.
- 4.** Your experimental questions should be your first consideration in choosing a factorial design because the number of questions you ask will determine how many IVs your experiment will have.
- 5.** measured
- 6.** Your friend has listed six IVs that she wishes to include in her experiment. This is too ambitious for a beginning project, and the results could be incredibly difficult to interpret because of the many potential interactions.
- 7.** a

CHECK YOUR PROGRESS PAGE 298

- 1.** Your experimental design has two IVs, classification and sex. Both are between-subject variables because a participant cannot be in more than one group. You would use the

factorial between-group design and test the data with a factorial ANOVA for independent groups. Your block diagram would look like the following:

	Male	Female
Freshman		
Sophomore		
Junior		
Senior		

- 2.** This experimental design has two IVs: test-taking practice and study course. Both are within-subject variables because each participant takes the tests repeatedly and takes the study course. You would use the factorial within-group design and use a factorial ANOVA for correlated groups to analyze the data. Your block diagram would be the following:

No study course			
With study course	1st Try	2nd Try	3rd Try

3. Again, you still have two IVs: the practice and the course. However, the study course is now a between-subjects variable because some participants take it and some do not. Test-taking practice is a within-subjects variable because each participant takes the test three times (repeated measures). Thus, you have factorial mixed-groups design and would use a factorial ANOVA for mixed groups. Your block diagram would be the same as the one shown for Question 2—only the assignment of participants to groups differs.

- 4.** An interaction effect is the simultaneous effect of two IVs in such a way that the effects of one IV depend on the particular level of the second IV. A significant interaction overrides any main effects because the interaction changes the meaning of the significant main effect.

5. a

6. This experiment has two IVs—the seasons and sex. Sex is always a between-subjects variable, so you must choose between the factorial between-groups design and the factorial mixed-groups design. Your decision rests upon how you wish to treat the participants with regard to the seasons. If you use different participants for each season, both IVs would be between subjects and you would use the factorial between-groups design. If you wish to use the same participants for each season, you would use the factorial mixed-groups design. Your selection of a particular design would probably depend on how many participants you can find—with large numbers, using a between-subjects manipulation is acceptable.

7. This experiment has two IVs—restaurant and age. Age must be between subjects; restaurant could be either between or within subjects based on whether the same or different people eat at each restaurant. Thus, your choices are the same as for Question 6.

Chapter 13

CHECK YOUR PROGRESS PAGE 308

1. random assignment; experimental design
2. It is essential to randomly assign participants to groups to help assure that the groups are equal before the experiment.
3. Random selection refers to choosing participants from a population in a nonsystematic manner. Random assignment refers to allocating those randomly selected participants to different groups in a nonsystematic manner.
4. It is problematic to use the pretest–posttest control group design to help with internal validity because this design cannot control all internal validity threats.
5. d
6. The drawback of using the Solomon four-group design to control for internal validity is that there is no statistical test to analyze all the data from this design.
7. The posttest-only control group design is a good choice for controlling internal validity because it seems to control the internal validity threats covered in Chapter 8. Drawing a general diagram of the posttest-only control group design is not possible because the design can vary depending on the number of treatment groups in the experiment (e.g., see Figures 13-5 and 13-6, p. 306).

CHECK YOUR PROGRESS PAGE 321

1. The use of single-case designs has decreased recently because of statistical innovations—an increase in both statistical tests and computerized analysis programs.
2. A single-case design can be used to disprove a theory because it takes only a single negative instance to invalidate a theory's general applicability.
3. baseline; three
4. a
5. For: Visual inspections of data may lead to incorrect conclusions. Statistical analysis gives increased accuracy.
Against: Treatments should produce effects that are visually apparent. Clinical significance is more important than statistical significance.
6. 1-C; 2-A; 3-B
7. You might be forced to use an A-B single-case design in the real world because it can be impractical or unethical to reverse a treatment. Many examples are possible. Suppose you work with an individual who suffers from anorexia or bulimia. You institute a treatment and the individual shows marked improvement. However, you fear that removing the treatment could result in a recurrence of the problem.

CHECK YOUR PROGRESS PAGE 331

1. Experimental designs involve manipulation of IVs, control of extraneous variables, and measurement of DVs so that cause-and-effect relationships can be determined. Ex post facto designs involve IVs that have already occurred or that are predetermined, such as

sex. Quasi-experimental designs involve manipulating IVs with groups of participants that are predetermined, such as causing something to vary between men and women.

- 2.** You might choose to use a quasi-experimental design rather than an experimental design in a situation when random assignment is impossible or if you wish to evaluate an ongoing program.
- 3.** 1-B, C, F; 2-A, D,* E

*Figure 13-15 (p. 327) gives the impression that the interrupted time-series design includes pretesting—in some cases, it may actually be pretesting. In many situations that use this design, however, the observations before the treatment are not true pretests because they were not made with an experiment in mind; rather, they may often simply represent data that were already available (e.g., sales records). This is a fine distinction, and it could easily be argued that both designs actually include pretesting.

- 4.** The key element in Geronimus's research that allowed a more positive conclusion about teen pregnancy was the realization that finding a true control group would be impossible; hence the need for a strong comparison group.
- 5.** We are more certain about our conclusion in the helmet law case because it simulated an A-B-A design, whereas the AFDC study was only an A-B design. If we can return to the baseline condition, we usually get a better reading of our results.
- 6.** d

Chapter 14

CHECK YOUR PROGRESS PAGE 370

- 1.** "APA format" refers to the accepted, standard form for preparing the results of psychological research that has been adopted by the American Psychological Association. APA format was developed in order to bring standardization and uniformity to the publication of research in the field of psychology.
- 2.** 1-C; 2-E; 3-B; 4-D; 5-F; 6-A
- 3.** The abstract is the most widely read section of most research reports because it is published in *Psychological Abstracts* and computerized indices such as PsycINFO.
- 4.** The introduction is similar to a typical term paper because it summarizes a body of knowledge about a relatively narrow topic. It is different from a term paper because it provides a rationale and lead-in to a particular experiment.
- 5.** c
- 6.** The method section is designed to describe (a) the research participants, (b) the equipment or materials used in the research, and (c) the manner in which the experiment was conducted. Although all this information is important, the procedures are probably the most important because they allow other researchers to replicate the study. If different participants or equipment (materials) were used, external validity would simply be increased.
- 7.** inferential; descriptive
- 8.** No, you could not use figures or tables as your sole information in a results section. Figures and tables are meant to supplement descriptive and inferential statistics, not to replace them.

- 9.** Although this answer involves opinion, you can make a strong argument that the discussion section is the most important section because it summarizes the experiment's evidence, ties that evidence into previous findings, and draws an overall conclusion.

10. The reference for a chapter from an edited book is more complex than a journal article or book reference because such a reference must include information about both the chapter and the book.

11. 1-E; 2-B; 3-C; 4-H; 5-A; 6-D; 7-F; 8-C

CHECK YOUR PROGRESS PAGE 386

1. You should not write your research paper in the style of a famous writer because your primary goal is communication rather than entertainment.
 2. Three strategies used to improve writing style are (a) write from an outline; (b) write a draft and put it away for some time before rereading it; and (c) have another person read and evaluate your writing. You will have to answer for yourself what you need to change about your writing in order to incorporate these suggestions.
 3. c
 4. a. Jones (1995) conducted an experiment.
 - b. The participants sat in desks around the room.
 - c. I (or "The experimenter") projected the stimulus items from the rear of the cubicle.
 - d. I found a significant interaction. or The data showed a significant interaction.
 5. a. The participants that were older were tested first. (It is important to distinguish the older participants from the younger.)
 - b. The room, which was a classroom, was used for testing. (The notation that the room is a classroom is probably merely a side comment that is not vital in describing the room.)
 6. a. Because you are the oldest, you should go first. "Since" was incorrect because a justification was implied rather than a time reference.
 - b. "Since" I began that class I have learned much about statistics. Correct because since denotes something has happened after beginning the class.
 - c. "While" we are watching TV, we can also study. Correct because while denoted that watching TV and studying will occur at the same time.
 - d. Although you are older than I, I should still go first. "While" was incorrect because a contrast was implied rather than a time reference.
 7. Orientals—Asians
mankind—humankind or humans
mothering—parenting
homosexuals—lesbian, gay
elderly—elderly people (or use specific age range)
girls and men—women and men
chairman—chair
depressives—depressed patients
 8. a + b = c trial-by-trial
-1 Enter: Your name

Name Index

A

Adair, J. G., 20
Altman, J., 108, 325
Anastasi, A., 79
Aronson, E., 37
Arruda, J., 237, 238

B

Ball, R., 106
Barlow, D. H., 310, 311, 316, 317, 320
Baumrind, D., 31
Beach, F., 162
Beauchamp, T. L., 30
Belquist, J. E., 375
Benjamin, L. T. Jr., 61
Bennett, S. K., 166
Benson, D., 68
Bhan, P., 270
Bickman, L., 322
Blum, T., 79
Blumenthal, A. L., 7
Boas, F., 50
Bodner, D., 79
Boice, M., 102, 106
Boring, E. G., 49, 165
Bowman, J. A., 15
Brewer, C., 204
Bridgeman, P. W., 101
Broad, W. J., 42
Broca, P., 54
Brooks, C., 60
Brownlow, S., 89, 102
Brunolli, S., 68
Buckner, C., 233, 234, 236, 237, 239
Burkley, E. III, 34, 80, 211
Burleson, J., 162
Burt, C., 43

C

Cacioppo, J. T., 6
Campbell, D. T., 147, 148, 149, 150,
151, 153, 154, 159, 160, 161, 162,
164, 168, 303, 304, 305, 317, 318,
322, 325, 326, 327, 329
Capaldi, J., 95, 96
Carlo, G., 128
Carlsmith, J. M., 37
Cass, K., 136, 334, 336
Cattell, J. M., 87
Childress, J. F., 30
Claus, M., 128
Cochran, C. D., 79
Coley, S., 114
Cook, T. D., 154, 322, 325, 326, 327,
329
Cooley, E., 272
Corbin, J., 54, 55

Corkin, S., 311
Cresswell, J., 50
Critchlow, B., 6

D

Darley, J., 94
Davis, S. F., 12, 15, 36, 162, 380
Dickson, A., 136, 161, 334, 336
Dickson, W. J., 62
Donovan, K., 87
Doyle, A. C., 1, 18, 53, 96, 131, 147,
157, 197, 204, 276, 311
Driggers, K. J., 2, 5, 7, 14, 21, 34
Dukes, W. F., 310, 311
Dunham, R., 242, 344, 346

E

Early, K., 281
Ebbinghaus, H., 162, 310
Echols, L., 68
Einstein, A., 16
Erickson, C. A., 15
Erikson, E., 165
Ervin, S., 102
Espiritu, G., 114

F

Featherston, J., 50, 52
Fechner, G., 310
Festinger, L., 5
Fine, M., 56
Fisher, R. A., 310
Fode, K. L., 134
Franklin, M., xii
Freeman, N., 89, 90
Freud, S., 165

G

Gaines, M., 219
Galton, F., 76
Garcia-Lopez, S., 278
Gargano, G., 102, 106
Genovese, K., 94
Geronomus, A. T., 324
Gibson, A., 36, 62, 63, 64
Gilbert, D. T., 142
Giuliano, T. A., 136, 334, 336
Glass, D. C., 72
Glesne, C., 52, 53
Goodall, J., 52
Goodwin, C. J., 38, 49, 61, 87
Gray, A., 71
Greenier, K., 237, 238
Grover, C. A., 15
Guba, E. G., 51, 52

Guerin, B., 19
Guthrie, R., 165

H

Hadsall, C., 21
Hall, G. S., 165, 166
Hall, R. V., 314, 315, 316
Harlow, H. F., 167
Harris, S., 68
Hatch, J. A., 50
Hayes, K. M., 36
Hedrick, T. E., 322, 323, 324, 327, 329
Helmreich, R., 73
Helms, T., 2, 5, 7, 14, 21, 34
Hersen, M., 310, 311, 313, 314, 316,
317, 320
Hilts, P. J., 311
Holloway, A., 281
Howard, R. W., 166
Huff, D., 45

I

Irving, L., 38, 144
Ittayem, N., 272

J

Jacobson, L., 134
James, N., 346
Jastrow, J., 7, 333
Jones, J. H., 30
Jones, M., 129, 166
Jones, R. R., 313

K

Kalat, J., 38
Kargi, E., 106
Kaser, R. T., 20
Kazdin, A. E., 313, 314
Kaziauskas, M., 344
Keith-Spiegel, P., 12
Kelland, M., 344
Kemp, D., 381–382
Keppel, G., 249
Kiker, K., 129
Kimmel, A. J., 29
Kimpel, J. D., 106
Koberlein, T., 68
Koestler, A., 16
Koger, S., 270
Korn, J. H., 41
Krantz, D. S., 72
Kuse, A. R., 325

L

Landrum, R. E., 12, 380
Lang, P. J., 318

Langley, W. M., 15

Larey, A., 115

Latané, B., 94

Lee, D. J., 166

Licht, C., 69, 70, 169,
383–385

Lincoln, Y. S., 51, 52

Lonner, W. J., 166

Losch, M. E., 6

Lucas, G., 270, 346

Lucas, J., 71

Luehring, J., 108, 325

M

Mahoney, M. J., 43

Malone, P. S., 142

Malpass, R., 166

Marin, G., 166

Marzola, L., 60

Matlinoski, B., 50

Matsumoto, D., 166

McClellan, K., 206, 207, 208, 209,
210, 211, 215, 218, 219, 263, 264,
274, 279

McFarland, S., 34

McInerny, R. M., 204

Mead, M., 50

Medewar, P. B., 15

Melamed, B. G., 318

Milgram, S., 29, 31, 38

Miller, H. B., 38

Miller, H. R., 36

Miller, J. G., 142

Miller, L. A., 15

Montgomery, M., 87

Mook, D., 167

Mooney, R. L., 73

Morris, J., 136, 334, 336

N

Nash, S., 19, 102

Nelson, S., 30

O

Orne, M. T., 161

P

Pepperberg, I. M., 15

Peshkin, A., 53

Prieto, L. R., 40, 41

Pritchard, M., 70, 346

Proctor, R., 95, 96

Prosser, I. B., 166

Proud, J., 270
Punzo, D., 89, 90

R

Rainville, B., 270
Reagan, T., 38
Reid, D. W., 72
Rewey, K. L., 12
Richard, M. M., 15
Riepe, S. M., 267, 268, 269, 270
Robinson, K., 67, 68
Roethlisberger, F. J., 62, 161
Rog, D. J., 322
Rolg, M., 60
Rosenberg, J., 270
Rosenthal, R., 34, 121, 132, 133,
134, 136
Rosnow, R., 121, 136
Ross, L., 142

S

Sagles, S., 114
Salthany, J., 60
Sasson, R., 30
Saufley, W. G. Jr., 249
Scall, R., 89
Schroeppel, N., 127
Sears, D. O., 164
Serrano-Rodrigues, J., 68
Shipman, S. L., 327
Shuter, R., 143
Singer, P., 38
Skinner, B. F., 16–17, 20, 310, 313
Slewert, S., 106
Small, W. S., 38
Smith, K., 36, 62, 63, 380
Smith, L., 219
Smith, R. A., 162
Snyder, M. L., 72
Solomon, L. Z., 169
Sou, E. K., 38, 144
Southwick, L. L., 6
Spatz, C., 208, 310
Stanley, J. C., 147, 148, 151, 159, 160,
161, 162, 164, 168, 303, 304, 305,
322, 327
Stapp, J., 73
Strauss, A., 54, 55
Sullivan, C., 233, 234, 236, 237, 239
Sumner, F. C., 165

T

Tan, 54
Tavris, C., 165

Templer, D. I., 72

Theis, J., 15

Thomas, J., 12

Titchener, E. B., 61

Tokunaga, H., 249

Torres, A., 36, 62, 63, 68

Traffanstedt, B., 319, 320

U

Urbina, S., 79

V

Vandenberg, S. G., 325
Vaughn, E., 218
Vaught, R. S. k., 313
Velayo, R., 114
Vohra, N., 20

W

Walker, K., 237, 238
Walker, W., 34
Walter, K., 102
Ware, E. E., 72
Watson, J., 61, 310
Weinberg, R. A., 219
Weinrott, M., 313
Wells, R., 212, 213, 215
Welterlin, A., 315
Welton, G., 281
Werner, H., xii
Wiederman, M. W., 12
William of Occam, 204
Williams, W. H., 38
Williamson, N., 102
Woods, E., 206, 207, 208, 209, 210,
211, 215, 218, 219, 263, 264,
274, 279
Wundt, W., 49, 61, 82, 310

Y

Yandell, L., 242, 344, 346
Yang, K. S., 143
Yates, C., 382–383
Young, J., 34

Z

Zahregian, P., 114
Zajonc, R., 19
Zapatel, J., 278
Zenner, C., 68, 70

Subject Index

A

A-B-A-B design, 316–17
A-B-A design, 315–16
A-B design, 314–15
Abscissa, 178
Abstracts, 336–38
Achievement tests, 78–79
Acknowledging participants, 375
ACT (American College Test), 66, 77, 148, 149, 151, 172, 173
Air Crib, 16
Alpha errors, 199
Alternative research designs, 300–330 internal validity and, 300 posttest-only control-group design, 305–7 pretest-posttest control group design, 304–5 quasi-experimental, 321–30 single-case experimental design, 308–20 Solomon four-group design, 305 A (measurement), 314 American Psychological Association (APA), 7, 8, 25 code of ethics, 30, 32–38 Ethical Standard 8 for the Conduct of Research, 32–33 Ethical Standard 8.09–Human Care/Use of Animals, 39 Ethical Standard 8.10–8.15— Reporting/Publishing, 46 American Psychological Society, 8 Analysis, 4 Analysis of variance (ANOVA), 244 for factorial designs, 284–85 one-way, 244, 248–56 rationale of, 245–47 repeated-measures, 244 two-way, 287–96 types of, 244 Analytic statements, 92 Animal Care and Utilization Committee, 29, 39 *Animal Learning & Behavior*, 162 Animals, ethical use of, 38–39 Animal Welfare Act of 1966, 39 ANOVA. *See* Analysis of variance (ANOVA) APA-format research report, 50, 333–85 abstracts, 336–38 APA editorial style, 376 apparatus subsection, 346 appendices, 362 author notes, 362–63 bias in language, 375 checklist for APA formatting, 377–79 discussion section, 353–56

headings, 195 introduction, 338–43 manuscript preparation, 376 materials subsections, 344–45 method section, 343–53 participants subsection, 343–44 procedure subsection, 346 reference section, 342 results section, 348–53 testing instrument(s) subsection, 346 title page, 334 writing guidelines, 371–73 Apparatus for experiments, 129–31 dependent variable recording and, 131 individual variables and, 129–30 Apparatus subsection, APA format, 346 Appendices, 362 Aptitude tests, 79 Archival and recorded data sources, 59–61 compared to experimental method, 61 potential problems, 60–61 Archives of the History of American Psychology, 60 Arousal reduction, 6 Association for Psychological Science (APS), 25 Asymptotic, 254 Attitude questionnaires, 149 Author notes, 362–63 Availability of participants, 127–28 Axial coding, 55

B

Back translation, 144 Balancing, 115–16 Bar graphs, 177–78 Baseline, 312 Baseline measurement, 312 Beta errors, 199 Between-groups variability, 215, 244, 245 Between-subjects comparison, 208 Between-subjects factorial designs. *See* Random assignment Bias language, 375 B (measurement), 314 Books, citations for, 360 chapters from edited books, 360–61 Bystander effect, 94

C

Carryover effects, 122–23 Case studies, 54 Case-study approach, 309 Cause-and-effect relations, 89–91, 300 internal validity and, 300

Central tendency, measures of, 173–76 choosing, 174–76 mean, 174 median, 174 mode, 174 Chapters from edited books, citation for, 360–61 Citations, 26, 342 for books, 360 chapters from edited books, 360–61 periodical articles, 359–60 for periodical articles, 359–60 of references, 45 for World Wide Web sources, 361 Classroom lectures, research ideas from, 19 Code of ethics American Psychological Association (APA), 30 Cognitive dissonance theory, 5 Cohorts, 82 College students as participants experimental validity and, 163–64 Comparative psychology, 162 Complementary information, APA format, 350–53 Complete counterbalancing, 119 Completely randomized ANOVA, 244 Completely randomized designs. *See* Random assignment to groups Completely within-groups (within-subjects) designs. *See* Non random assignment to groups Computer statistical output interpretation of, 248–56, 287–96 Concurrent validity, 77 Conditional matrices, 56 Confirmability of data, 51, 87 Confounded experiments, 105, 208–9, 235 internal validity and, 300 Confounders, 103–5. *See also* Extraneous variables (confounders) Confounding, 105, 300 Constancy, 114–15 Content validity, 77 Contradictory statements, 92 Control, 87–88 Control groups, 206 Control of extraneous variables balancing, 115–16 constancy, 114–15 counterbalancing, 116–24 elimination, 114 randomization, 113 Control procedures, 235 random assignment as, 235 Control techniques, 113–24

- Convenience sampling, 163–64
 Correctness, 106
 Correlated assignment to groups, 209–13
 matched pairs/sets, 210, 236,
 268, 269
 natural pairs/sets, 212–13, 237, 270
 repeated measures, 211–12,
 236–37, 269–70, 311–12
 within-subjects comparison, 212
 Correlated groups, 235, 268
 Correlated-groups designs, 215–17
 Correlated samples, 235–38. *See also*
 Non-random assignment to groups
 one-way ANOVA for, 253–56
 two-way ANOVA for, 291–93
 Correlation, 188–92
 correlation coefficient, 188–91
 perfect positive, 190
 zero correlation, 188
 Correlational studies, 3, 66–68
 correlational research, 67–68
 nature of correlations, 66–67
 Correlation coefficient, 188
 Pearson product-moment, 191–92
 Counterbalancing, 116–24
 carryover effects, 122–23
 complete, 119
 differential carryover, 123–24
 incomplete, 119–21
 sequence or order effects, 121–22
 within-group, 117–19
 within-subject, 117
 Credibility of data, 51
 Criterion validity, 77
 Cross-cultural psychology, 141–42, 166
 cultural immersion, 53
 effect on research, 143
 ethnocentrism, 142
 experimental hypothesis and, 143
 fundamental attribution error and, 142
 knowledge, truth and, 142
 methodology and analysis, 143–45
 methodology and analysis issues,
 143–45
 participants/sampling procedures,
 143–44
 research and, 141–45, 143
 research effects, 143
 survey or questionnaire type and, 144
 variable selection and recording, 143
 Cross-sectional research, 82
 Cultural response set, 145
- D**
- Data
 confirmability of, 51
 credibility of, 51
 demographic, 74
 dependability of, 51
 fabrication of, 43–44
 qualitative, 51–52
 transferability of, 52
 Debriefing sessions, 37–38
 Deception, 34, 37–38
 Decisions, 4
 Deductive logic, 94–95
 Definitions, operational, 101
 Degree or amount, 106
 Degrees of freedom, 195, 216–17
- Demand characteristics, 135–36,
 138, 161
 Demographic data, 74
 Denatonium saccharide, 15
 Dependability of data, 51
 Dependent variables (DV), 3, 89, 105–8,
 204–5
 characteristics of, 108
 recording more than one, 107–8
 recording or measuring, 106–7, 131
 reliability of, 108
 selection of, 106
 two-group design and, 204
 validity of, 108
 Descriptive research methods, 3,
 59–66
 archival and recorded data sources,
 59–61
 choosing behaviors and recording
 techniques, 63–66
 naturalistic observation, 61–63
 reactance or reactivity effect, 62
 Descriptive statistics, 171–86, 350
 Design. *See* Experimental design
 Differential carryover, 123–24
 Diffusion (imitation of treatment),
 154, 305
 Directional research hypotheses, 96
 Discussion section, APA format, 353–56
 comparing results to previous
 research, 353–56
 interpreting results, 356
 restating results, 353
 Double-blind experiments, 138
 DV. *See* Dependent variables (DV)
- E**
- Economy of expression in writing, 372
 Editorial style, APA, 376
 Effect size, 200
 Elimination, 114
 Emic, 142
 Empirical measurements, 86–87
 Environmental generalization, 158
 Environmental variables, 102
 Errors
 fundamental attribution, 142
 Type I (alpha), 199, 289
 Type II (beta), 199
 Error variability, 215, 245, 285
 in factorial design, 284–85
 Ethics in psychological research
 animals, use of, 38–39
 at end of research, 42–47
 of experimenters, 40
 Institutional Review Boards (IRB), 29,
 39–40
 need for, 29–32
 obligations of researcher, 40
 of participants, 40–42
 Ethnocentrism, 142, 166
 Ethnographic inquiry, 52–53
 Etic, 142
Even the Rat Was White (Guthrie), 165
 Everyday occurrences, research ideas
 from, 17–18
 Expectancies, experimenter, 133–34,
 134–35
 Experience independent variables, 102
- Experimental analysis of behavior,
 310–11
 Experimental design, 203–19, 231–42,
 259–81
 alternative designs, 300–30
 factorial design, 260–72
 factorial versus two-group and
 multiple-group designs, 273–76
 multiple-group design, 232–38
 multiple-group design, variations on,
 241–42
 multiple-group versus two-group,
 238–40
 multiple-independent-groups versus
 multiple-correlated-groups,
 240–41
 statistics and, 220–21
 Experimental groups, 206
 Experimental hypotheses, 6, 91–97.
See also Research hypotheses
 culture and, 143
 Experimental mortality, 305
 Experimenters
 characteristics of, 133
 controlling effects of, 134–35
 expectations of, 133–34, 134–35
 as extraneous variables, 132–35
 Hawthorne effect and, 62
 physiological and psychological
 effects of, 134
 reactance or reactivity effect and, 62
 responsibilities of, 40
 Rosenthal effects and, 134
 Experiments, 3, 88–89
 confounded, 105
 true, 219
 Ex post facto research, 3, 69–70, 219,
 242, 278
 External validity, 157–69
 culture and, 166
 demand characteristics and, 161
 generalization and, 157
 methodological threats to, 159–62
 multiple-treatment interference, 162
 need for, 167–69
 participant threats to, 162–66
 race and, 165–66
 reactive arrangements, 161
 Extraneous variables (confounders), 89,
 103–5, 113, 206, 300
 control of, 112–24. *See also* Control
 of extraneous variables
 experimenters as, 132–35
 participants as, 135–39
- F**
- Fabrication of data, 43–44
 Factorial design, 260–72
 assignment to groups, 268
 comparing different amounts of
 an IV, 277–78
 control issues in, 276–77
 error variability, 284–85
 experimental questions, 276
 interpreting statistics in, 287–96
 mixed assignment to groups,
 271–72
 with more than two IVs, 278–81
 naming, 283–84

- nonrandom assignment to groups, 269–71
number of groups or levels, 261–68
number of independent variables, 260–61
post hoc comparisons, 296
practical considerations, 277
random assignment to groups, 269
rationale of factorial ANOVA, 284–85
selection of, 276–77
statistical analysis of, 283–84
three-way, 279–81
treatment variability, 284–85
versus two-group and multiple-group designs, 273–76
using measured IVs, 278
variations on, 277–81
- F**
Factors, 259. *See also* Independent Variables (IV)
Failure to replicate, 18
Falsifiability, principle of, 93
Figures, 350. *See also* graphs
Findings
 significance of, 7
Focus groups, 53–54
Follow-up tests (post hoc comparisons), 251, 296
Forced alternative questions, 72
Freedom, degrees of, 195
Frequency polygons, 178
Fundamental attribution error, 142
- G**
General implication form, 92–93
Generalizability. *See also*
 Transferability
 participants and, 127
Generalization, 157–59
 environmental, 158
 population, 158
 temporal, 158
General Social Survey (GSS), 60
Good participant effect, 136–37
Grammatical guidelines, 373–75
Graphs
 bar graphs, 177
 frequency polygons, 178
 histograms, 177
 line graphs, 178–81
 pie charts, 176–77
Grounded theory, 54–56
 axial coding in, 54
 conditional matrices in, 56
 open coding in, 54
 process models in, 55
 selective coding in, 54
 transactional systems in, 55
Groups
 assignment to, 207–13, 234–38, 268–72
 experimental *versus* control, 206–7
- H**
Hawthorne effect, 62
Headings for APA-format research, 195, 334, 363–64
Level 1 headings, 336, 364
Level 2 headings, 364
- Level 3 headings, 343, 364
Level 4 headings, 363, 364
Level 5 headings, 364
Heterogeneity of variance, 223
Histograms, 177
History, 147–48, 304
History effects, 78
History of Experimental Psychology, A, 49
Homogeneity of variance, 223
How to Lie with Statistics (Huff), 45
Human subjects
 debriefing sessions and, 37–38
 deception and, 34, 37–38
 ethical principles related to, 32–34
 informed consent, 29, 34–36
 participants at minimal risk, 36
 participants at risk, 36
 right to withdraw, 34–35
 vulnerable populations and, 36–37
Human Subjects Review Panel, 29, 39
Hypotheses, 4, 6. *See also* Experimental hypotheses; Research hypotheses
Hypothesis testing
 new view of, 95–97
- I**
Idea presentation, 372
Ideas, research, 14–19
Imitation of treatment (diffusion), 154
Incomplete counterbalancing, 119–21
“In Defense of External Invalidity” (Mook), 167
Independent groups, 208, 217, 235, 268
Independent-groups designs
 advantages of, 217
Independent samples, 235. *See also*
 Random assignment
 one-way ANOVA for, 248–53
 two-way ANOVA for, 287–91
Independent variables (IV), 3, 69, 88, 204–5, 232, 259
 apparatus and, 129–30
 experience, 102
 measured, 218–19, 278
 more than two, 278–81
 multiple-group designs and, 232–33
 participant, 102–3
 physiological, 102
 presentation of, 129–30
 stimulus or environmental, 102–3
 two-group design and, 204
 types of, 102–3
Index term selection, 20–21
Inductive logic, 94
Inferential statistics, 171, 192–99
 null hypothesis, 192
 one-tail *versus* two-tail tests, 196
 significance and, 192
 significance testing, logic of, 196–99
 t tests, 193–95
Informed consent, 29, 34–36
Inspiration, research ideas from, 16
Institutional Review Boards (IRB), 29, 39–40
Instrumentation, 305
Instrumentation and automation, 135
Instrumentation (instrument decay), 150, 305
- Interactions, 264
 in factorial designs, 285–87
 with selection, 153–54, 304
of selection and treatment, 160–61
synergistic effects, 286
of testing and treatment, 159–60
Interlibrary loans, 25
Internal validity, 300
 alternative research designs and, 300
 cause-and-effect relations and, 300
 confounding and, 300
 diffusion or imitation of treatment, 154
 experimental design and, 304–7
 history and, 147–48
 instrumentation (instrument decay) and, 150
 interactions with selection and, 153–54
 maturation and, 148
 mortality and, 152–53
 nonreactive measures, 149–50
 practice effect and, 148–49
 protecting, 155, 303–7
 random assignment and, 303–4
 reactive measures and, 149
 research design and, 303–7
 selection and, 151
 statistical regression and, 150–51
 testing and, 148–50
 threats to, 147–54
Internet searches, 21
 evaluating resources, 24–25^{t2}–2
 literature reviews and, 21–23
Interobserver reliability, 65
Interrater reliability, 77
Interrupted time-series design, 327–30
Interval scale, 172
Interviews
 personal, 75–76
 telephone, 76
Introduction, APA format, 338–43
Inventories, 3, 76–79
IV. *See* Independent variables (IV)
- J**
Journal of Applied Behavior Analysis, 311, 313
Journal of Experimental Psychology: Animal Behavior Processes, 162
Journal of Psychological Inquiry, 8, 10
Journal of Psychology and the Behavioral Sciences, The, 9
Journal of the Experimental Analysis of Behavior, 311
Just noticeable difference (jnd), 310
- K**
KISS principle, 277
Knowledge, acquisition of, 3–4
Knowledge acquisition, 3–4
- L**
Labels, 375
Language bias, 375
 labels, 375
 level of specificity, 375

Latency or duration, 106
 Level 1 headings, 336, 364
 Level 2 headings, 364
 Level 3 headings, 343, 364
 Level 4 headings, 363, 364
 Level 5 headings, 364
 Levels, 206, 233
 Likert-type scales, 72
 Line graphs, 178–81
 Literature reviews, 4, 5
 computerized research, 21–23
 conducting, 20–28
 index term selection, 20–21
 integrating results of, 25–28
 relevant publications, obtaining, 23–25
Logic
 deductive, 94–95
 inductive, 94
 knowledge and, 3–4
Longitudinal research projects, 82
Lying with statistics, 45

M

Mail surveys, 74–75
 Main effects, 264
 Manuscript page header, 334
 Manuscript preparation, 376–79
 Marginal significance, 288–89
 Matched pairs/sets, 210, 236, 268, 269
 Matching variables, 235, 236
 Materials subsections, 344–45
 Maturational, 148, 304
 Mean, 174
 regression to, 150–51
 Mean squares, 249
 Measured independent variables, 278
 Measurement, 172
 baseline, 312
 central tendency, measure of, 173–76
 of dependent variables, 106–7
 graphing, 176–81
 scales of, 172–73
 variability, measures of, 182–86
 Measure of central tendency, 171–72, 173–76
 Median, 174
 Method section, APA format, 343–53
 apparatus subsection, 346
 materials subsection, 344–45
 participants subsection, 343–45
 procedure subsection, 346–48
 testing instrument(s) subsection, 346
 Milgram's obedience studies, 29, 31
 Mixed assignment (mixed groups), 268, 271–72
 Mixed factorial designs, 271
 Mixed samples
 two-way ANOVA for, 293–96
 Mode, 174
 Mortality, experimental, 152–53
 Multiple-choice questions, 72
 computer-generated output, 248–56
 Multiple-group designs, 232–57
 assigning participants to groups, 234–35
 comparing, 240–41

control issues in, 240
 independent variables and, 232–33
 independent versus correlated groups, 235
 interpretation of statistics, 247–56
 number of levels in, 233
 placebo effect, 241
 practical considerations, 240–41
 principle of parsimony, 233
 statistical analysis of, 243–47
 treatment groups in, 233–34
 versus two-group designs, 238–40
 variations, 241–42
 Multiple independent variables, 278
 Multiple observers, 65
 Multiple-treatment interference, 162

N

Narrative record, 64
 National Health Research Act (1974), 29
 National Opinion Research Center, 60
 Naturalistic observation, 52, 61–63
 Natural pairs/sets, 212–13, 237, 270
 Nay-sayers, 137
 Nazi experiments, 29–30
 Negative correlation, 66, 228
 Nominal scale, 172
 Nondirectional research hypotheses, 96
 Nonequivalent group design, 323–27
 Nonexperimental research, 49, 59–83
 Non-random assignment to groups, 209–13, 235–38, 268, 269–71
 Nonreactive measures, 149–50
 Nonsystematic sources of research ideas, 16
 Normal distribution, 184–85
 bell curve, 184
 Nuisance variables, 109–11
 Null hypothesis, 192
 Nuremberg Code, 30
 Nuremberg War Tribunal, 30

O

Objectivity, 86–87
 Observational studies
 choosing behaviors and recording techniques, 63–66
 interobserver reliability, 65
 situation sampling, 64
 time sampling, 64
 Observers
 Hawthorne effect and, 62
 interobserver reliability, 65
 as participants, 53
 reactance or reactivity effect of, 62
 Occam's (Ockham's) razor, 204
 One-way ANOVA, 244
 for correlated samples, 253–56
 for independent samples, 248–53
 Online databases
 www.apa.org, 21
 Open coding, 55
 Open-ended questions, 73
 Operational definitions, 101, 244
 Oral presentation, 382–83
 Ordinal scale, 172
 Ordinate (Y axis), 178

P

Parsimony, principle of, 204, 233, 277
 Participant observation, 52
 Participants
 acknowledging, 375
 availability of, 127–28
 characteristics as IVs, 102–3
 characteristics of, 103
 college students as, 163–64
 controlling effects of, 138–39
 cross-cultural sampling procedures and, 143–44
 debriefing of, 37–38
 demand characteristics, 135–36, 138
 experimenter's expectations and, 133–34
 as extraneous variables, 135–39
 generalizability and, 127
 good participant effect, 136–37
 likelihood of success and, 127
 at minimal risk, 36
 number of, 128–29
 as observer, 53
 power of statistical tests and, 129
 precedent and, 127
 responsibilities of, 40–41
 at risk, 36
 types of, 126–28
 types of research project and, 128
 vulnerable populations, 36–37
 white rats as, 162–63
 women as, 164–65
 Participants at minimal risk, 36
 Participants at risk, 36
 Participants subsection, APA format, 343–44
 Participatory action research (PAR), 56–57
 Passive voice, 373–74
 Past research, ideas from, 18–19
 Pearson product-moment correlation coefficient, 191–92
Pedagogical Seminary (Journal of Genetic Psychology), 165
 Perfect positive correlation, 190
 Periodical articles, citations for, 359–60
 Personal interviews, 75–76
 Personality tests or inventories, 79
 Physiological and psychological effects, 134
 Physiological independent variables, 102
 Pie charts, 176–77
 Pilot testing, 71
 Placebo effect, 241
 Plagiarism, 42–43, 356
 Population generalization, 158
 Populations, 79
 Positive correlation, 66, 226
 Poster presentation, 381–82
 Post hoc comparisons, 251, 296
 Posttest-only control-group design, 305–7
 Power, 129
 Practice effect, 148–49
 Precedent, 127
 Precision and clarity in writing, 372
 Presence-absence manipulation, 218

- Presentations, 4, 7–10, 383–84
 Pretest–posttest control group design, 304–5
 Principle of falsifiability, 93
 Principle of parsimony, 204, 233, 277
 Problem definition, 4
 Procedure subsection, APA format, 346–47
 Process, 55
 Professional activities, 379
 oral presentations, 382–83
 poster presentation, 381–82
 presentations, 4, 7–10
 publication, 4, 7–10, 383–84
 student participation in, 379–80
 Programmatic research, 296
Psi Chi Journal of Undergraduate Research, 8, 10, 60
Psi Chi (National Honor Society in Psychology), 8, 9
 PsychFIRST, 21
 PsychINFO database, 21, 23
Psychology: Health, Happiness, and Success, 61
 Publication, 4, 7–10, 383–84
Publication Manual of the American Psychological Association, 7, 291, 334
 “Publish or perish,” pressure to, 44
Psychological Abstracts, 21
- Q**
- Qualitative research, 3, 49–57
 case studies, 54
 characteristics of, 50–51
 data analysis of, 51–52
 ethnographic inquiry, 52–53
 focus groups, 53–54
 grounded theory, 54–56
 methods of, 52–54
 naturalistic observation, 52
 overview, 50
 participant observation, 52
 participatory action research (PAR), 56–57
 Quantitative methods, 64–65
 Quasi-experimental designs, 321–30
 history of, 322
 interrupted time-series design, 327–30
 nonequivalent group design, 323–27
 representative designs, 323–30
 uses of, 322
 Questionnaires, 3, 70–76
 Questions, types of, 72–73
- R**
- Racism, 165–66
 Random assignment, 207–8, 235, 268, 303–4
 Random assignment to groups, 207–9, 268–69
 correlated samples (nonrandom assignment to groups), 235–38
 multiple-groups design and, 235
 Randomization, 113–14
 Random samples, 79
 Random sampling
 with replacement, 79
 without replacement, 79
 Random selection, 208, 304
 Range, 182
 Rate or frequency, 106
 Ratio scale, 173
 Rats as participants, 162–63
 Reactance/reactivity effect, 62
 Reactive arrangements, 161
 Reactive measures, 149
 Reasoning, types of, 94–95
 Recording techniques, choosing, 63
 References, 26, 342
 citation of, 45–47
 References, 356–62
 Reference section, APA format, 342, 356–62
 Regression to the mean, 150–51
 Reliability, 77, 108
 of dependent variables, 108
 split-half technique, 78
 test-retest procedure and, 78
 Repeated measures, 211–12, 236–37, 269–70, 311–12
 single-case experimental designs and, 311–12
 Repeated-measures ANOVA, 244
 Replication, 18, 87, 168
 Replication with extension, 168
 Reports, 7
 Reprints, 25
 Research
 Research, cross-cultural, 141–45
 Research design, 6. *See also* Experimental design
 Research hypotheses, 91–97
 analytic statements, 92
 characteristics of, 92–95
 contradictory statements, 92
 directional *versus* nondirectional, 96–97
 general implication form, 92–93
 principle of falsifiability, 93–94
 synthetic statements, 92
 Research ideas, 14–19
 characteristics of good, 14–16
 sources of, 16–19
 Research methods, 3
 importance of, 11–12
 Research process, 4–11
 analysis, 6–7
 components of, 4
 deception in, 34, 37–38
 design of, 6
 experiment, conducting, 6
 hypotheses in, 6
 informed consent, 34–36
 interpretation of, 7
 literature reviews, 5
 planning of, 4
 presentations and publication, 7–10
 problem definition, 5
 reporting of, 4, 7
 theories and, 5–6
 Research questions, developing, 19–20
 Research strategies, 81–83
 cohorts, 82
 cross-sectional, 82
 longitudinal research projects, 82
 single-strata approach, 82
 Response bias, 137–38
 nay-sayers, 137
 response sets, 137–38
 yea-saying, 137
 Response set, 137, 139
 Responsibilities
 of experimenters, 40
 of participants, 40–42
 Results section, APA format, 348–53
 Right to withdraw, 34–35
 Robustness, 223
 Rosenthal effects, 134
 Running head, 334
- S**
- Samples, 79
 Sampling
 populations, 80
 random, without replacement, 80
 random, with replacement, 80
 random samples, 80
 stratified random, 81
 SAT (Scholastic Aptitude Test), 66, 77, 148, 149, 151
 Scales of measurement, 172–73
 interval, 172–73
 nominal, 172
 ordinal, 172
 ratio, 173
 Scientific method
 components of, 86–88
 control, 87–88
 empirical measurements, 86
 objectivity, 86
 replication of findings, 87
 self-correction, 87
 Selection, 151, 304
 interactions with, 153–54
 internal validity and, 151
 Selective coding, 55
 Selective deposits, 61
 Self-correction, 87
 Sensory thresholds, 310
 Sequence or order effects, 121–22
 Serendipity, research ideas from, 16–17
 Sexism, 164–65
 Significance tests
 of findings, 7
 logic of, 196–99
 one-tail *versus* two-tail tests of, 196
 t tests, 193–95, 222–25, 225–28
 Single-blind experiments, 135
 Single-case experimental design, 308–20
 A-B-A-B design, 316–17
 A-B-A design, 315
 A-B design, 314
 baseline measurement, 312
 case-study approach, 309
 changing one variable at a time, 312–13
 experimental analysis of behavior, 310–11
 general procedures for, 311–13
 history of, 310
 repeated measures, 311–12

Single-case experimental design (*continued*)
 statistical analysis and, 313–14
 uses of, 310–11
 Single-strata approach, 82
 Situation sampling, 64
 Smoothness of expression in writing, 372
 Social facilitation theory, 19
 Society for Research in Child Development (SRCD), 37
 Society for the Experimental Analysis of Behavior, 311
 Solomon four-group design, 305
 Source tables, 249, 288
 Split-half technique, 78
 Standard deviation, 183–86
 Statements, types of, 92–94
 Statistical regression, 150–51, 304
 Statistics, 171
 calculating and computing, 181–82
 descriptive. *See Descriptive statistics*
 experimental design and, 220–21
 factorial designs, 283–84
 inferential. *See Inferential statistics*
 interpretation of, 222–28, 247–56, 287–96
 lying with, 45
 multiple-group design analysis, 243–44
 Pearson product-moment correlation coefficient, 191–92
 power and, 129
 significance testing of. *See significance tests*
 single-case experimental design and, 313
 translating into words, 224–25, 227–28, 252–53, 255–56, 289–90, 289–91, 293, 295–96
 Stimulus independent variables, 102
 Stratified random sampling, 81
 Success, likelihood of, 15–16
 Sum of squares, 249
 Surveys, 3, 70–76
 analytic, 71
 descriptive, 70
 development of, 71–73
 mail, 74–75
 pilot testing of, 71
 Synergistic effects, 286
 Synthetic statements, 92

T

Tables, 350–51
 Telephone interviews, 76
 Temporal generalization, 158
 Temporal meaning, words with, 374–75
 Testability, 14–15
 Testing, 148–50, 304
 Testing instrument(s) subsection, APA format, 346
 Test-retest procedure, 78
 Tests, 3, 76–79, 77
 achievement tests, 78–79
 aptitude tests, 79

characteristics of, 77
 personality tests or inventories, 79
 reliability of, 77–78
 significance tests. *See Significance tests*
 types of, 78–79
 validity, 77
That vs. which, use of, 374
 Theoretical considerations, 4
 Theories, properties of, 5–6
 Theory, research ideas from, 19
Thesaurus of Psychological Index Terms (APA), 21, 22
 Thesis statement, 338
 Three-way design, 279–81
 Time sampling, 64
 Title page, APA format, 334
 Transactional systems, 55
 Transferability, 52
 Treatment conditions, 206
 Treatment groups, 233
 Treatment variability, 285
 in factorial design, 284–85
 True experiments, 219
 Trustworthiness of data, 51
t tests, 193–95
 for correlated samples, 225–28
 for independent samples, 222–23
 Tukey HSD (honestly significant difference), 251
 Tuskegee syphilis project, 29, 30
 Two-group designs, 204–19
 advantages of, 215–17
 analysis of, 221
 between-groups variability, 215
 calculating statistics, 221
 choosing, 214–15
 comparing, 214–17
 control groups, 206
 control issues with, 215
 control of extraneous variables, 206
 degrees of freedom in, 216–17
 dependent variables and, 204
 error variability, 215
 experimental groups, 206
 independent variables and, 204
 interpretation of statistics, 222–28
 levels or treatment conditions and, 206
 versus multiple-group designs, 238–40
 principle of parsimony, 204
 statistical analysis of, 220–21
 statistical issues with, 215
 variations on, 218–19
 Two-way ANOVA
 for correlated samples, 291–93
 for independent samples, 287–91
 for mixed samples, 293–96
 Type I (alpha) errors, 199, 289
 Type II (beta) errors, 199

U

Unbiased language, 343

V

Validity, 77, 108
 concurrent, 77
 content validity, 77
 criterion validity, 77
 of dependent variables, 108
 external. *See External validity*
 internal. *See Internal validity*
 interrater reliability, 77
 Variability, 172, 215
 measures of, 182–86
 Variables
 dependent variables (DV).
 See Dependent variables (DV)
 extraneous variables (confounders).
 See Extraneous variables (confounders)
 independent variables (IV).
 See Independent variables (IV)
 nature of, 100–101
 nuisance variables. *See Nuisance variables*
 operational definitions of, 101
 Variance, 182–83, 250
 Vulnerable populations, 36–37

W

Web sites
 scholar.google.com, 21
 www.apa.org, 21
 www.apa.org/psycinfo/psycfirst.html, 21
 www.ipcpsr.umc.edu/gss/home.htm, 60
 www.srcd.org/about.html#standards, 37

Which vs. that, use of, 374
 White rats as participants, 162–63
 Willowbrook hepatitis project, 29, 30
 Within-group counterbalancing, 117–19
 Within-group variability, 128, 245.
 See also Error variability
 Within-subject comparison, 212
 Within-subject counterbalancing, 117
 Women as participants, 164–65
 World Wide Web sources, citation for, 361
 Writing guidelines, APA format, 371–73
 economy of expression, 372
 grammar, 373–75
 orderly presentation of ideas, 372
 precision and clarity, 372
 smoothness of expression, 372
 strategies to improve writing style, 372–73

Y

Yea-saying, 137, 139
 Yes–no questions, 72

Z

Zero correlations, 67, 188