

Becoming an  
Evidence-Based  
Practitioner

**PRACTICAL RESEARCH  
METHODS FOR EDUCATORS**

**Ennio Cipani**

# Practical Research Methods for Educators

*Becoming an Evidence-Based  
Practitioner*

**ENNIO CIPANI, PhD**

  
**SPRINGER PUBLISHING COMPANY**  
New York

**Ennio Cipani, PhD**, a graduate of Florida State University, is a licensed psychologist and a full professor in the Department of Special Education at National University–Fresno. He has published numerous articles, chapters, books, and instructional material (including software) in the areas of child learning and behavior management in homes and schools. Of particular note are *Punishment on Trial: A Resource Guide to Child Discipline* (2004) and *Functional Behavioral Assessment, Diagnosis, and Treatment: A Complete System for Education and Mental Health Settings* (Cipani & Shock, 2007, Springer Publishing). He has given many workshops at state and national conventions and has been doing in-home and school-based behavioral consultation for families and school personnel since 1982. He has dealt with many families and a variety of behavior problems, conducting assessment and intervention activities in homes and classrooms.

Copyright © 2009 Springer Publishing Company, LLC

All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the publisher or authorization through payment of the appropriate fees to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-646-8600, info@copyright.com or on the web at www.copyright.com.

Springer Publishing Company, LLC  
11 West 42nd Street  
New York, NY 10036  
www.springerpub.com

*Acquisitions Editor: Philip Laughlin*  
*Project Manager: Julia Rosen*  
*Cover design: David Levy*  
*Composition: Apex CoVantage, LLC*

Ebook ISBN: 978-0-8261-2236-0

09 10 11 12 / 5 4 3 2 1

The author and the publisher of this Work have made every effort to use sources believed to be reliable to provide information that is accurate and compatible with the standards generally accepted at the time of publication. The author and publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance on, the information contained in this book. The publisher has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

---

### Library of Congress Cataloging-in-Publication Data

Cipani, Ennio.

Practical research methods for educators : becoming  
an evidence-based practitioner / Ennio Cipani.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-8261-2235-3 (alk. paper)

1. Education—Research—Methodology—Textbooks. I. Title.

LB1028.C545 2009

370.7'2—dc22

2009000380

---

Printed in the United States of America by Hamilton Printing

# Contents

*Preface* vii

- 1 Science and the Scientific Method 1**
  - What Is Science? 1
  - Science and the Study of Variables 3
  - Descriptive Research Studies 6
  - Correlation Research Studies 11
  - Experimental Research Studies 20
  
- 2 Measuring the Dependent Variable 35**
  - Methodological Requirements for Measuring Student Behavior 35
  - Steps to Measuring Student Behavior 38
  - Measuring Student Learning 53
  - Summary 56
  
- 3 Basics of Single-Case Designs 59**
  - Characteristics of Single-Case Research Designs 60
  - Requisites for Collecting Single-Case Research Data 66
  - Determining an Experimental Effect 72
  - Group Versus Single-Case Experimental Designs 75
  - Summary 79
  
- 4 Single-Case Designs 81**
  - AB and Additive Designs 81
  - Reversal Designs 93
  - Multiple Baseline Designs 104
  - Multielement Designs (aka Alternating Treatments Designs) 114
  - Changing-Criterion Designs 121

<b>5</b>	<b>Types of Single-Case Research Studies</b>	<b>127</b>
	Deciding on the Type of Research Study	127
	Potential Thesis/Dissertation Research Topics	135
	Daily Report Card (DRC)	135
	Reinforcement Programs With a Response Cost (RC)	141
	Tap Me on the Shoulder!: Teaching Appropriate Attention-Getting Behavior	147
	Self-Monitoring of Appropriate or Inappropriate Behavior	151
	Sit and Watch (Time Out)	155
	Response Cards	160
	Appendix A: Reliability Formulas, Interobserver Agreement	167
	Appendix B: Raw Frequency Data for Figure 3.1	169
	Appendix C: Graphing Data: General Rules for the X, Y Coordinate Plane	171
	Appendix D: Raw Frequency Data for Figure 3.3	173
	<b>Index</b>	<b>175</b>

# Preface

It would be nice if the material you learned in any course could be demonstrated to be of importance in a future career. I am sure you have been through courses where certain topics that were covered seemed useful, relevant, and functional for your career in education. Of course the converse is also true. There are many topics that seem to have little relevance to what you will be required to know as a teacher in a special day class for elementary students, as a school administrator, as a junior high school math teacher, or in a variety of other positions on the front line of education.

If you are a future elementary school educator, you are probably not required to take a course in tax law or a course in advanced Boolean algebra (although some prospective high school math teachers may be required to take such a course). The relevance of such a course to elementary school teachers would certainly be questionable. It would not serve a functional purpose for the position requirements (although such courses may be of interest to some people).

You are probably taking this course because it is a requirement for a graduate degree that you are seeking. But you may feel that its function for your future career is limited at best and nonexistent at worst. I do not believe that the utility of developing research skills in practitioners is often questioned. The inclusion of a research course in a preparation program for educators is seen as a given. Unfortunately, the skills deemed necessary and desirable often have little practical utility for frontline classroom personnel. For example, many traditional texts on educational research methods written for prospective educational personnel provide extensive treatments of experimental designs that often prove impractical for classroom teachers and school administrators. Should we really expect that teachers will be able to carry out a two by two ( $2 \times 2$ ) group factorial design with random assignment, to determine whether extra phonics practice produces better reading on weekly reading test scores?

If this type of design or methodology is not realistic for practicing teachers, why would a text spend extensive time teaching it, as well as teaching the related statistical tests to analyze that data? While competency with such types of research methodologies and statistical analyses is a skill that an educational researcher working for the Educational Testing Service (ETS) would need to master, this would be overkill (and irrelevant) for current and future classroom practitioners.

I believe this text offers a fresh perspective on developing research skills for educators whose careers will be in the classroom. I believe this text is unique in that it identifies the functional requirements for conducting pragmatic research for everyday instructional personnel. Such a functional approach for future educators is desperately needed. I contend that the development of research skills should focus on two broad objectives for educational personnel who are (or will be) direct line service providers: (1) being producers of applied research that is relevant for classroom practice; and (2) being wise consumers and critics of the research findings reported in literature and everyday conversation.

## **BEING PRODUCERS OF APPLIED RESEARCH, RELEVANT FOR CLASSROOM PRACTICE**

Not all research is created equal. Some research studies provide answers and solutions to questions and problems relevant to everyday educational practice. Some do not. The distinction between relevant and irrelevant to everyday practice is often seen as the distinction between applied research and basic research. *Applied research* involves testing techniques and procedures that have an effect on classroom behaviors and learning immediately, in a real-life situation. For example, a second-grade teacher would like to use a strategy that helps her students who are just learning the English language to read more words with accuracy and fluency. If a number of teachers have also had difficulty with the reading skills of these students, a need is created for research to address this problem empirically. The findings of such research would be immediately useful to this teacher as well as to teachers who face the same type of problem with English as a Second Language Learners (ELLs).

This teacher is also enrolled in a graduate program in education and feels that this might be a great topic for a thesis. She reviews the research literature and discovers that there are many empirical investigations of this topic. Unfortunately, the overwhelming majority report no

experimental data to lend credibility to their assertions. This seems like a great topic for a thesis. After defending her proposal, she initiates the data collection on the dependent variable. She examines each student's scores on weekly 1-minute tests of accuracy and fluency. She collects such data on each student for 10 weeks. During this 10-week period, the ELL students had a 15-minute silent reading time as a strategy to increase accuracy and fluency with grade-level reading materials. Subsequently, upon viewing the results, the teacher verifies the need for improvement across many of the students whose reading accuracy and fluency is below grade level.

What intervention will this teacher use to produce (hopefully) greater reading accuracy and fluency in her ELL students? She has read about peer tutoring and wonders whether such an approach would work in her class. Several studies have demonstrated that peer tutoring is effective in raising oral reading scores, and it seems that using students who read well as peer tutors would be effective. But only the data from such a study would be proof positive. The teacher implements peer tutoring for the last 15 minutes of the class, taking the place of the 15-minute silent-reading period. While some of the articles she had read discussed which approach works best for ELL students, her study will provide evidence. Practitioners of science are concerned with results, not debate. This teacher, being an evidence-based practitioner, will implement and evaluate the efficacy of peer tutoring in her class with each of her ELL students, following the initial experimental condition (i.e., silent reading).

The results over the next 10 weeks confirm the teacher's suspicions. Having her ELL students spend 15 minutes on peer tutoring each day produced a greater number of correct words read per minute (than in the prior 10 weeks) for 80% of the students. Debate is one thing; evidence is another. As you can see, this research study has immediate relevance for this teacher; she learned that 15 minutes of peer tutoring was worth the time invested. While peer tutoring was found to be effective for these students, silent reading was not as effective. The teacher's decision to change the lesson format for the reading period to include peer tutoring is on solid ground. This teacher's applied research answers a question of immediate concern to her. It further validates the teacher's continued use of this strategy as having a solid efficacy base and may provide another study demonstrating the efficacy of peer tutoring with oral reading skills. Just as important, it adds to the research literature on peer tutoring by demonstrating a positive result with ELL students.

In the divergent approach, termed *basic research*, investigators address questions that may not be of immediate utility in real-life situations. Basic research is seen by many as a longitudinal process that may yield application after years of effort have been spent. The intent of the basic researcher is to develop some theory that may explain some aspect of student behavior and/or learning (Gay, Mills, & Airasian, 2006). For example, a basic research study may want to know if the ELL students have more difficulty reading accurately and fluently when they have certain deficits in verbal reasoning. This researcher may have a theory about verbal reasoning as a developmental requisite to reading. Her research will attempt to provide evidence for that theory. She subsequently identifies those students with lower scores on reading measures, as well as those students who have grade-level performance. She then measures their verbal reasoning via one of the commercially available subtests on intelligence measurement. The results illustrate that such students do have lower scores on the verbal reasoning subtest than students who do not have difficulty reading grade-level material in a fluent fashion. But what is the utility of such a finding for the classroom? Get students with higher verbal reasoning scores to enroll in your class? Basic research may or may not lead to application. In the educational arena, the track record of published basic research leading to useful findings is not stellar.

Too often, educational research is not seen as helpful to everyday practitioners. There is a simple explanation for this. Very often, it does not help classroom teachers. This lack of relevance can often create problems for graduate students who intend to conduct their thesis or dissertation in a school district where they are outsiders. Asking a teacher to participate in a research study that does not aid that teacher in educating or managing her classroom more effectively will make it harder to enlist the teacher as a participant. In contrast, asking teachers who have difficulty in getting students to transition more smoothly from one instructional activity to another to participate in a study that will address this problem area would probably be more welcome. The latter study will provide a direct benefit to the teacher(s) if one of the strategies being tested actually works!

Research with applied relevance can also address how to change the behavior of students. A hypothetical elementary grade school has difficulty with students coming in from lunch recess in a timely manner. Would a study that examines student attitudes about school and recess directly solve that problem? No! While such a study might have some information on how students feel about school, the result of conducting

this study will not be a change in student tardiness in coming in from lunch recess. In contrast, an applied study might examine several strategies aimed at getting students to line up within 30 seconds of hearing the recess whistle. Suppose the current level of students who line up within 30 seconds is 35%–70% across a 4-week period. If the school tries one strategy, for example, reminding students each day to get in their respective lines within 30 seconds of the recess whistle, then its result on student behavior can be compared against the prior level. If this strategy does not result in a significant improvement, another, more potent technique might be added, for example, loss of recess time for those students not in line at 30 seconds. If the level of students getting in line in a timely fashion now ranges from 60% to 90%, a management problem during recess is solved. With a few additional research design requirements, this data can demonstrate that such an approach, that is, daily reminders and recess fines, results in improved student behavior in this context. Again, the utility of such a finding is immediate, that is, student behavior is improved.

Applied research in education, to be relevant, needs to focus more on empirical tests of interventions and strategies that purportedly produce changes in student behavior and learning. This is particularly evident for those students whose behavior is seen as resistant to typical efforts. What is not helpful in developing a technology of classroom behavior change is conducting a study to survey these students' opinions about their self-esteem. No behavior changes! Nor do we have any evidence that a treatment that could develop increased levels of self-esteem would produce a change in real-life behavior! If the researcher conducting such a study believes that enhancing self-esteem will have an effect on student behavior, prove that directly! Show that an intervention program that produces changes in students' self-esteem results in a change in their test scores, without any confounding interpretation. We need fewer studies that show associations and more studies that show treatment effects directly.

## **BEING WISE CONSUMERS AND CRITICS OF RESEARCH FINDINGS**

A physician who treats persons with various heart problems relies on research, in the form of clinical trials, to determine which treatment regimens to administer to her patients. She must keep abreast of the research findings with respect to the statins (a drug class for treating

high cholesterol) that have been found to be effective with patients and present the fewest risks in terms of side effects. Her use of these drugs is based on a rigorous research program demonstrating that such a drug will lower cholesterol levels in patients with such medical presenting problems. A practicing physician does not rely on evidence that does not demonstrate an effect on the criterion variable, in this case, the cholesterol levels of patients. Medical personnel have been evidence-based practitioners for years. Theories are not evidence! But how does this relate to the use of teaching and management approaches in classrooms?

As someone who is about to enter (or has entered) the profession of education, how often have you heard someone use “the research says . . .” to bolster his or her argument for implementing some procedure, strategy, or policy? Very often, such statements are not challenged. Yet it may be questionable whether there is research that demonstrates the efficacy of such a procedure. Here is a great example.

*Is there really a memory pill?* There are many products on the market claiming that they improve memory. They often claim that there is clinical research demonstrating that memory is improved. But the devil is in the details. Here are some hypothetical research studies claiming that the results support the daily use of a particular memory pill as a treatment that improves people’s memory:

- Charts, developed by scientists, that show how this pill affects neurotransmitter substances in the brain
- MRI scans that show increased blood flow in the brain after a group of participants have been using the memory pill for 6 weeks
- Correlations between blood flow in the brain and short-term memory (measured by tests)
- Data from participants taking the pill reporting that they felt it improved their memory
- Testimonials from actors about their use of the memory pill, reporting an improved ability to remember their lines
- Results demonstrating that gains in memory (as measured by a test) were attained by participants using the pill for a prescribed period of time and that such gains were not obtained with a group of participants who were given a placebo pill (where the members of neither group were aware of which pill they were getting)

Which of the above studies constitute empirical evidence regarding the efficacy of a memory pill in improving the memory? None! Did you fall for the charts on neurotransmitters and the information on blood flow as evidence that memory pills produce a better memory? Such studies would require a *leap of faith* to accept the theory that increasing blood flow to the brain will result in better memory. It did not demonstrate that such a result actually occurred. It showed that the pill produces changes in blood flow (that is verified) but failed to take the next step, which means demonstrating that such a result in the brain leads to a change in memory ability (by actually measuring the change).

*Brain research has to be credible!* What about a study demonstrating that people who have greater blood flow in the brain do better on memory tests than people who show less blood flow. Does the study show that a product that can increase blood flow will improve memory? No, it does not! It may be the case that increasing blood flow will result in better test results, but that part of the experiment was not conducted! Again it requires a leap of faith. To demonstrate the treatment efficacy of the pill, it would have required that the researchers follow up such a study with a clinical trials study involving people with poor memories (and presumably less blood flow). If the researchers then demonstrate, with those participants, that increasing the blood flow results in changes in memory, then such a study provides empirical evidence.

As a practitioner in education, you will often be confronted by people who say “the research says . . .” After this course, you will respond with, “Show me the data!” You will be able to critically evaluate whether the conclusion they are drawing is warranted. If they can produce experimental data, demonstrating that some variable causes a change in student behavior, you will see that their conclusions are warranted. If their research findings require a leap of faith, you will now be able to spot that. You can inform them that they do not have the necessary evidence regarding the efficacy of their theory or approach.

*Don't be fooled!* Suppose a group of “experts” develop a theory that the most effective manner of developing reading skills in young children is to use a multisensory method and approach. They reason that if children listen to stories (providing auditory sensation), then draw pictures about the stories (providing a visual sensory component), and finally mime the story for a peer, reading scores will improve. These experts get testimonials from reading teachers who agree that such an innovative strategy for teaching reading decoding and comprehension would no doubt work. The experts conduct a survey and find that 90%

of the reading specialists surveyed scored the approach as either certainly efficacious or probably efficacious. Does this constitute empirical evidence that this approach is effective? Only if scientific knowledge is acquired by conducting Gallup polls! The leap of faith is that the results of a future study would parallel the survey data. In contrast, readers of this book will not accept such a leap of faith but require studies demonstrating that the approach under consideration results in better reading test scores than other approaches.

## OVERVIEW OF THE TEXT

This text begins by presenting the scientific method in chapter 1. The reader is given a basic discussion of the measurement of two specific variables: the dependent variable and the independent variable. The chapter covers three basic forms of research: descriptive, correlation, and experimental. The chapter will target the skills necessary for the practitioner to become a wise consumer of research findings. The remaining chapters will then build the skills needed to conduct applied research in classrooms.

Chapter 2 provides detail on the requisites for single-case research and on its methodological designs. Three characteristics of single-case designs are delineated. Chapter 3 presents information on measuring the dependent variable in single-case research studies. The sine qua non of measurement in applied research is a direct measurement of behavior. Issues of reliability and validity are presented next. The rest of the chapter then focuses on steps to measuring behavior and learning as the primary dependent measures used in single-case research designs.

Chapter 4 presents a variety of single-case designs for use in classroom research projects. For each design, the chapter presents a brief description of the design, some figures illustrating the research design features with hypothetical data, and some research studies that have utilized such a design. Each type of single-case design is discussed with regard to its advantages and limitations in demonstrating cause-and-effect relationships.

Chapter 5 provides a discussion of the four types of applied research that can be conducted: demonstration, comparative, parametric, and component analysis. The remainder of this chapter provides a variety of possible examples of applied research with respect to a number of behavioral and instructional interventions and looks at how the four types

of studies could be conducted. Chapter 5 presents a discussion of types of applied research and a host of potential research areas.

## REFERENCE

Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational research: Competencies for analysis and applications* (8th ed.). Columbus, OH: Merrill/Prentice Hall.

# 1

## Science and the Scientific Method

### **WHAT IS SCIENCE?**

Science is the systematic study of phenomena to gain knowledge about the world. Scientists have two core sets of beliefs: (1) a belief in determinism, and (2) a belief in empiricism rather than speculation as a means of gaining knowledge (Cooper, Heron, & Heward, 2007). This chapter reviews the basic premises of science and shows how different research methodologies provide the mechanism for collecting different types of knowledge.

### **Belief in Determinism**

Do you believe that phenomena occur randomly (accidentalism) or that events are the result of a combination of occurrences leading to a result (determinism; Cooper et al., 2007)? Scientists generally subscribe to determinism. For example, if you believe in determinism, you will view changes in the weather as a function of changes in certain weather events. Your explanation for hurricanes and their strength (Categories 1–5) will rely on an analysis of certain weather and oceanic variables that set the conditions under which hurricanes form. The degree to which meteorologists better understand how such variables impact weather

makes them more capable of predicting weather patterns and hurricane strength during the summer months in the Caribbean and the Atlantic Ocean.

If you do not believe that a combination of certain levels of relevant variables affects weather conditions, then you will see events such as hurricanes as unpredictable. This approach might best be called “weather happens.” You will be inclined to believe that the weather on one day will have no effect on the next day’s weather. Given our society’s sophistication with explanations of what variables influence weather patterns, a deterministic view is very plausible in understanding weather. However, centuries ago, the explanation offered for weather events relied on some aspect of a deity.

In education, determinism as a philosophical position views environmental and inheritable characteristics as variables that explain behavior. Of course, to educators, environmental variables are more relevant. Environmental variables can be studied and potentially altered to produce desired changes in the rate of behavior and level of academic performance. Determinism in education can be characterized by the adage “Teaching (method) matters!” If learning were a function of random events, then teaching would become irrelevant. One would just wait for “it” (i.e., student learning) to happen. When we are lucky, students learn. When we are unlucky, they do not. In contrast, a belief that what we do as teachers affects learners (in both desired and undesired directions) requires a belief that learner behavior is determined in some part or in large part by environmental conditions.

## **Belief in Empiricism**

Where is the evidence? For scientists, knowing something (knowledge) requires reliable and valid evidence. Consider this statement: This child’s behavior is driven by an innate desire to be good. The statement speculates that the child’s behavior is a function of an innate desire. But how would you measure any child’s innate desire? Further, how would a contention such as the one presented above be proven? If you were able to derive a reliable and valid method of measuring a child’s innate desire, you would then have to demonstrate this relationship. Subsequently, you would have to show that children who have this innate desire behave in one manner and children who do not possess such a desire behave in another (in some quantitative manner). If this sounds a lot like “The id made me do it!” you now see that explanations of human behavior have

evolved from speculative to empirical in the last half century (at least among the majority of professional people).

Empiricism is based on an objective, quantitative measurement of a phenomenon of interest (Cooper et al., 2007). Such a measurement must be demonstrated to be free from the bias of the researchers (Cooper et al., 2007). Scientific findings obtained via an empirical approach are neither good nor bad but value free. Demonstrating that a nuclear reaction occurs when nuclear particles are made to collide with each other is a fact of science. That fact is neither positive nor negative but a matter of empirical demonstration. How people use such knowledge becomes a social value issue.

In an empirical approach, experimentation is the vehicle for gaining knowledge about the world. We learn about a phenomenon by subjecting it to study. Further, to develop a knowledge base, a number of studies related to the phenomenon must be conducted (Sidman, 1960). Replication of results with additional studies is essential because one finding alone does not make a law of nature! If an empirical study is truly devoid of the personal bias of the experimenter, other researchers should be able to replicate the findings, as long as the variables are measured in the same manner as they were in the original study. Experimentation may demonstrate either that two variables covary or that one variable can cause changes in the other.

## SCIENCE AND THE STUDY OF VARIABLES

### Questions to Ponder

- If you were interested in knowing whether students learn better with cooperative learning, what other condition(s) would have to be tested?
- What is the relationship between the dependent variable and the independent variable?

### What Is a Variable?

A variable is a phenomenon that varies along some quantitative measure (Matheson, Bruce, & Beauchamp, 1978). For example, if you are told that your child's weight can vary on a given day by up to 2 pounds, depending on the time of day, this means that your child's weight is variable across time. However, if you are told that his or her weight at 3:20 p.m. was

112 pounds, then that measure is not variable (rather, it is a constant). In other words, a variable varies in quantity over a number of circumstances. Examine the statements in the left hand column of Table 1.1 to determine whether a variable or a constant is implied.

## Two Major Variables: One Depends on the Other

Scientific inquiry is concerned with two major variables: the dependent variable and the independent variable. The *dependent variable* exemplifies its name: Its value depends on another variable (the independent variable). Remember in basic algebra class you were often asked to plot a linear equation, such as  $y = 3x + 5$ , on an  $x, y$  coordinate graph. If I ask you, “What is the value of  $y$  in that equation?” what would be your answer? “It depends!” And you would be right! The value of  $y$  depends on the value you assign to  $x$ . If  $x$  is given a value of 2, what is the value of  $y$ ? Eleven. We can change the value of  $y$  simply by changing the value of  $x$ . When  $x$  is given a value of 3, the value of  $y$  becomes 14, as the value of  $y$  varies with the change in the value assigned to  $x$ . The value of  $y$  depends on the assigned value of  $x$ . In contrast, the equation  $y = 3$  does not vary; hence it is not a variable but a constant.

Table 1.1

### WHICH OF THE FOLLOWING?

STATEMENT	VARIABLE (YES/NO)
The length of your right foot at 9:00 on May 25, 2009	No, one value
The shoe sizes of residents of Cocoa Beach, Florida	Yes, many values (Cocoa Beach has more than one resident)
The number of times you scold your dog each day for a 1-month period (let's assume you have a dog)	Yes
The square footage of your house/apartment	No, exact amount
The time it takes you to run 100 meters in four consecutive races	Yes, variation across four different races
The length of time you spend texting your friends each day for 1 week	Yes

In applied educational research, the dependent variables of interest are concerned with student behavior (often called the *criterion variable* in relationship studies). In this context, the study of student behavior refers to any measure of student behavior or performance. Examples of student behavior can be any of the following: the number of student disruptive incidents in a cafeteria period, the number of positive comments made by one student to another, or the number of plastic bottles recycled during lunch periods. Student performance can be measured in any of the following ways: test scores, the number of items correct on assignments, or the length and sophistication of expository writing samples.

The *independent variable* is the phenomenon that the researcher manipulates, and whose effect on student behavior or performance the researcher studies. In education, four main types of independent variables often studied in experimental research are the following: (1) type of learning materials, (2) instructional method, (3) reinforcement contingencies, and (4) length of instruction/treatment (Gay, Mills, & Airasian, 2006). Suppose a researcher is interested in studying the effects of several different learning materials on student learning. Since the independent variable must vary, one has to specify at least two conditions (or levels) of the learning-materials variable. It is insufficient to simply implement one method and study the effect on student performance. Why? Because the independent variable does not vary in that circumstance: it is constant. However, if you have two sets of learning materials, say, Math Curriculum A and Math Curriculum B, you now have variation. Consequently, you can now study how that variation in the independent variable affects student performance. In other words, what happens to student performance when Curriculum A is in effect versus when Math Curriculum B is in effect? Does student performance vary significantly as a function of which of these two sets of materials is in place? This study would be given the following title: “The effect of two different types of curriculum material on student math scores measuring 10 instructional objectives.”

A research study will often delineate in its title the two variables studied. Table 1.2 gives the titles of two hypothetical research studies. The dependent variable is underlined, while the independent variable is italicized.

In the first study, the independent variable, number of rehearsal items with feedback, will vary at least over two levels. One level could be 3 student rehearsal items with feedback per lesson, while another

Table 1.2

**DEPICTION OF DEPENDENT AND INDEPENDENT VARIABLES**

1. *The effects of number of rehearsal items with feedback on 11th-grade students' acquisition of the skill of computing correct values for two variables in a linear system*
2. *The effects of rule reminders on student compliance with classroom rules during teacher lecture in a 2nd-year Spanish class*

level could be 10 items with feedback per level. Both these conditions will be evaluated against the 11th-grade students' performance on math items measuring the solving of unknowns in a linear system. In the second study, the effects of rule reminders on the level of student compliance with classroom rules will be assessed. This will again require that the independent variable will have at least two levels, for example, daily reminders from the teacher about the classroom rules versus weekly lectures about the rules.

Scientific inquiry has three basic purposes, from which are derived three basic methods of science. The three basic purposes are as follows: (1) describe the phenomenon; (2) examine relationships between variables; and (3) determine cause-and-effect relationships (via manipulating experimentally relevant variables). The types of research studies for each of these purposes are descriptive research, correlation research, and experimental research, respectively. The remainder of this chapter will present each of the methods that are used to uncover such information.

**DESCRIPTIVE RESEARCH STUDIES****Questions to Ponder**

- Can a descriptive study determine what causes student behavior? Why or why not?
- Would a descriptive study be indicated if one wanted to determine how many second-grade students can memorize a 3-minute song in one 30-minute training session?
- What are the differences between descriptive observational research and survey research?

- Why is it not acceptable to make generalizations about actual student behavior from survey research?
- Why is a descriptive observational research study unable to reach conclusions on cause-and-effect relationships?

## Descriptive Observational Research

Description is perhaps the easiest form of gaining knowledge and a first step toward subsequent scientific inquiry. You simply watch a phenomenon and describe what you observe. This is termed descriptive observational research. Suppose you wanted to learn more about ants. If you are a scientist, you would have to come in contact with ants in order to directly study and observe their patterns of behavior. It would be inappropriate to speculate about ants because you have had experience with frogs, birds, or other animals. You would first identify where ants live, to allow for direct observation. You would then possibly observe these ants for some period of time. You would begin to look for patterns of ant behavior and begin to come up with hypotheses about ant activities. A descriptive study would then possibly measure the frequency or duration of certain patterns of ant behaviors in relation to other events. In contrast to other methods of science, you simply describe the behavior of the ants without altering their environment in any way.

In descriptive research, observation of the phenomenon is conducted by the researcher. It is essential to obtain a recording of the observation(s) that allows for quantification. A quantifiable method of collecting information from observations of behavior separates a more scientific approach from a simple anecdotal record. If you have a hypothesis about the existence of a certain phenomenon in an educational setting, your first step would be to collect information that describes that phenomenon. A descriptive study would then measure that phenomenon, in a quantitative manner. Such data would then allow you to determine whether your hypothesis is accurate.

What would be an example of a descriptive observational study with children? Suppose an educational researcher speculates that children who engage in disruptive behavior during the lunch recess are often not disciplined. This researcher translates the area of inquiry into the following research question: How often are consequences (i.e., discipline) deployed for children who engage in disruptive behavior during lunch recess? This research question can be addressed by collecting data on two phenomena. First, the researcher observes the number of times children engage

in disruptive behavior during lunch recess. Second, the researcher also measures how often consequences (delineated a priori) are levied shortly after each disruptive incident. If the rate of consequences following an act of disruptive behavior is 10% (meaning that 90% of the time no consequences are delivered following the act), this researcher's hypothesis is confirmed. Table 1.3 provides hypothetical data on the percentage of times disruptive behavior results in the utilization of at least one of three identified consequences over a 5-day period.

This hypothetical study merely describes how often consequences (from the list of three measured by the researcher) befall acts of disruptive behavior during lunch recess. The percentage of times that consequences were deployed ranged from a low of 20% to a high of 50% of the times when disruptive behavior occurred. This study cannot address any question related to the effects of one variable (consequences or lack thereof) on the other (rate of disruptive behavior). Descriptive observational research simply describes conditions (in a quantitative fashion) as they currently exist.

A study conducted by a group of researchers at Valdosta State University sought to determine whether the rate of correct responses was different during initial instruction (i.e., new material) than during practice lessons (Gunter, Reffel, Barnett, Lee, & Patrick, 2004). Three observers collected data on correct response rates in 104 elementary classrooms from seven schools. During 5-minute observations in a given classroom, the observer marked whether the academic question was taken from new instructional material (initial instruction) or from previously reviewed

Table 1.3

**DESCRIPTIVE ANALYSIS DATA**

	DATE				
	MARCH 2	MARCH 3	MARCH 4	MARCH 5	MARCH 6
Frequency of acts of disruptive behavior	10	20	12	15	10
Frequency a consequence followed an act of disruptive behavior	4	10	4	3	4
Percentage of times consequences were deployed	40%	50%	33%	20%	40%

material (practice lesson). Correct responses were then marked for each individual student answer. In order for correct responses to be recorded when a group choral response was made, 80% of the students' answers had to be correct. The median correct response rate for initial instruction was 2.8, while the review or practice lessons resulted in a median correct response rate of 4.1 (which was statistically significant). While such data reveal naturalistic rates of such behavior across many classrooms, they do not provide the effect of this variable (initial versus review lessons) on student acquisition of material (e.g., probe test data). The mechanism lacking in descriptive observational studies that precludes causal analysis is the failure to actively manipulate one of the variables, that is, the independent variable. This translates to an inability to determine cause-and-effect relationships, irrespective of how compelling the obtained observational data are.

## Survey Research

Survey data often falls into the category of descriptive research. Survey research gathers the opinions of the survey respondents and presents these data in a summary form. Conducting a survey study that examines how many students with attention deficit hyperactivity disorder (ADHD) ingest above-average levels of sugar would be an example of a descriptive study using survey data. Each student with ADHD participating in the study would possibly fill out a questionnaire about his or her sugar intake. The survey could have the students check off the category that best describes his or her sugar intake. This category system allows the researcher to determine for each respondent whether daily sugar intake is below, at, or above the usual levels of intake.

The results would merely report the percentage of students sampled whose intake of sugar exceeded some designated level (in the category of above-average intake). The finding might be that 50% of the students with ADHD sampled reported daily sugar intake that was above that consumed by other students. When such data are collected by surveying the intended participants (either verbally or using paper and pencil), it is termed survey research.

Survey research contrasts with descriptive observational research in that the criterion variable is measured indirectly. The sugar intake study with students with ADHD could be turned into an observational study by changing the data collection method. If the researcher collects data by observing these students and measuring their sugar intake directly, it becomes a descriptive observational study. Hypothetically, the researcher

may observe these students at lunch time, recording whether they eat a candy bar or drink soda for lunch. She may then report that 70% of her sample of students with ADHD ate up to one candy bar during lunch (allowing for not finishing it), while 35% of those observed ate more than one candy bar at lunch. Note that such a study does not identify what effect such activities may have (although the author may want you to take a leap of faith and infer what the effect would be). It simply describes a phenomenon.

## Conducting a Descriptive Research Study

Conducting a descriptive observational study is relatively straightforward in terms of the data collection process (see Table 1.4). Direct observation of the phenomenon is required. Selection of the variable(s) to measure via observation is an early decision in the research process.

### Utility of Descriptive Studies for Educational Practitioners

Descriptive studies that entail the collection of survey data do not contribute significantly to identifying effective intervention techniques for behavioral and educational learning problems. First, the accuracy and reliability of self-report survey data is very suspect. Can one trust that what people say in a survey reflects what their response would be a week from the time they filled out the questionnaire?

Table 1.4

#### STEPS TO CONDUCTING A DESCRIPTIVE OBSERVATIONAL STUDY

1. Delineate the variable(s) you are interested in, and how these will be measured.
2. Define the group of students who will be involved in the data collection process and the setting for the study.
3. Measure the student or environmental variable in a quantitative fashion across the students from the designated setting.
4. Collect the data for a period of time to achieve some stable measure of the variable being measured.
5. Report the data in some usable form, for example, frequency, percentage, and so forth.

Second, the assumption that opinion data provide insight into effective strategies is flawed. Survey data reporting that 85% of dentists prefer a brand of toothpaste with a licorice taste do not translate to a finding that licorice is an effective component in toothpaste. Consensus, even at high levels, does not provide scientific evidence that something works under certain conditions. Similarly, finding that most teachers who are judged to be great teachers prefer to use mild discipline approaches when dealing with behavior problems does not validate such approaches. *A Gallup poll, even with a sample of people judged to be great at their craft, does not equate to a science of cause-and-effect with respect to student behavior.* Could it be that great teachers may have great students? In that case, how can the finding help teachers who have less-than-great students?

Descriptive observational studies can have some value in initially exploring some classroom phenomena. However, descriptive observational research does not allow for an analysis of cause-and-effect. Describing a phenomenon, while possibly useful in determining the extent to which a phenomenon occurs, does not provide any direction for solving any observed problem(s). Cause-and-effect analysis can only be derived from a research methodology that actively manipulates the independent variable and assesses the effect on the dependent variable.

## CORRELATION RESEARCH STUDIES

### Questions to Ponder

- Why do correlation studies not allow for causal conclusions? What would a study have to demonstrate for one to conclude that a difference in teaching strategy caused changes in student learning?
- Explain why a correlation coefficient of  $-0.77$  indicates a strong relationship between two variables. Given this negative correlation, would high scores on the predictor variable translate to high or low scores on the criterion variable?
- What would be the danger of inferring cause from a correlation study showing that eating more peanut butter is statistically related to better student performance in the first two periods of the school day? In other words, what would be the danger if, as a result of this correlation finding, schools start asking all students to start eating peanut butter in the morning?

- What does an inverse correlation coefficient illustrate between two variables,  $a$  and  $b$ ? Describe what happens to the values on variable  $a$  when the values on variable  $b$  decrease.
- Does a correlation coefficient of .90 depict a cause-and-effect relationship between two variables? Why or why not?

## Examining the Relationship Between Two Variables

In educational and psychological research, studies often measure two or more variables, without any experimental manipulation, and attempt to determine whether there is a relationship (or association) between them. A relationship exists between two variables when knowing the value of one variable allows you to estimate, to some degree, what the value would be of the other variable. Here is a simple, more concrete example using two variables, labeled  $x$  and  $y$ . We begin by measuring these two variables across 10 circumstances. We measure the variable called  $x$  across these 10 circumstances. We also measure the variable called  $y$  across these same 10 circumstances. We find that the value of  $y$  is always two times the value of  $x$  (remember basic algebraic equations:  $y = 2x$ ). When  $x$  is 4, the value of  $y$  is 8. When  $x$  is 12,  $y$  is 24. In that scenario, we would have a direct linear relationship and perfect prediction. If someone tells you the value of  $x$  (called the *predictor variable*), you can predict what the value of  $y$  (called the *criterion variable*) is. While research on human behavior is not perfect in regard to prediction, educational and psychological researchers try to discern which variables correlate with student variables such as achievement, performance, and/or behavior. Provided the two variables are quantified, you can study the extent of the relationship between them.

When two events covary with high probability, prediction becomes possible. If height predicts weight to some degree, one could give a reasonable estimate of weight given height. For example, if someone is over 6 feet 2 inches tall, would you guess that she or he would weigh less than 120 pounds? Probably not. And 999 out of 1,000 times you would be right.

Suppose we hypothesize that getting to school on time is a predictor of school success. You select a number of children to observe and from whom to obtain data on how often they are late to school over a period of time. Some children get to school on time every day. Other children may be late 1 or 2 days a week. Maybe some other children are late 4 or 5 days a week. Based on attendance data that is collected across 20 weeks

within the academic year, each student is identified as belonging to one of three categories: (1) frequently late, (2) sometimes late, or (3) infrequently late. Students are classified as frequently late if the mean number of days late per week is between 3.0 and 5. Students whose mean number of days of being late to school is between 1.0 and 2.9 are classified as sometimes late. Finally, students whose mean number of days of being late to school is less than 1.0 are classified as infrequently late. Once you have gathered such data, you would then determine whether the level of getting to school on time is associated with school success. For the purposes of this hypothetical example, we could define school success as each student's test score on a standardized reading achievement test.

Now that we have quantifiable measures on the two variables of interest, we can determine how related these variables are. Let's say the results depict the following: Students who are in the infrequently late category are also the students with the highest reading comprehension test scores. Concurrently, those who are in the frequently late category generally have the lowest scores. Given these data, the results illustrate that the two variables covary. Knowing something about a student in terms of school arrival times allows you to estimate (with some degree of accuracy) the student's reading achievement test score. If I told you there is a student entering the third grade who is late 4 to 5 days per week, what would you guess is his or her reading score? You would guess it is low. If the strength of the correlation between these two variables is very high, your guess has a very good chance of being correct.

## **Correlation Research Addresses Degree of Association**

Correlation research addresses two questions (Cohen, Manion, & Morrison, 2003, p. 193). First, is there a relationship between two (or more) variables of interest? Second, what is the direction and what is the magnitude of the relationship? Correlation research involves proving that some variable covaries with a more complex variable (the dependent variable).

Correlation studies are frequently used to develop or build theories about child behavior and performance. Here is a hypothetical example. A researcher believes (i.e., has a theory) that getting plenty of sleep is related to better school performance for children with ADHD. In attempting to gather evidence that such a theory may be accurate,

she decides to conduct a study testing the relationship between the two variables.

This researcher specifies the population of interest as children with ADHD. She obtains a sample of participants who are hopefully representative of this population (children with ADHD), as her theory must have broad application to children with ADHD in general. She selects 120 high school students with ADHD, from 25 school districts, as the participants for her study. She defines school performance as each student's test scores on two midterm tests. This allows her to measure school performance in a quantitative manner, for example, using midterm scores in spelling and math (two dependent variables). To measure sleep levels, she develops a survey. She has the students fill out this survey, giving information about the amount of sleep they get over a 3-week period (self-report data). If sleep length is related to performance (as measured by weekly test grades), then one would see higher test scores for those students who report more hours of sleep each night. Students who score poorly on the tests would be the ones who report fewer hours of sleep per night. As you can see (Table 1.5), generally, the more hours of sleep a student reports he or she gets, the higher the test scores. Conversely, the fewer hours of sleep a student reports he or she gets, the lower the test scores in spelling and math (first nine participants depicted).

Table 1.5

**RAW DATA FOR TEST SCORES AND AVERAGE NUMBER OF HOURS SLEEP PER NIGHT**

<b>PARTICIPANT NO.</b>	<b>AVERAGE NO. HOURS SLEEP/DAY</b>	<b>TEST GRADE MATH</b>	<b>TEST GRADE SPELLING</b>
1	8.25	82%	90%
2	4.75	50%	35%
3	6.0	50%	50%
4	7.75	76%	81%
5	3.75	40%	60%
6	9.0	95%	88%
7	6.0	70%	60%
8	8.5	100%	100%
9	5.5	60%	70%

## Magnitude of Association: The Correlation Coefficient

In this hypothetical researcher's sample, it seems that higher levels of self-reported sleep covary with test performance. But what is the magnitude of the relationship in the above study? Is sleep highly correlated with test grades, or is the relationship fairly weak? The determination of this is quantitative. You simply do not look at the raw data sets and proclaim a strong relationship. Rather, correlation research expresses the magnitude of the relationship between the two variables as a correlation coefficient.

A correlation coefficient ranges from  $-1.0$  to  $+1.0$ . Correlations of  $1.0$  and  $-1.0$  are perfect and could be expressed by a linear equation ( $y = 3x + 17$ , or  $y = -3x + 17$ ). Knowing the value of  $x$  (number of hours of sleep) yields only one  $y$  value (test score on math or reading). Correlations of less than  $1.0$ , or greater than  $-1.0$ , demonstrate a relationship that is less than perfect. The further away from these two endpoints the correlation coefficient is, the weaker the relationship between the two variables. When an association (relationship) is strong between two variables, the correlation coefficient is closer to  $1.0$  or to  $-1.0$ . A correlation coefficient of  $.78$  demonstrates a stronger relationship between the two variables than a coefficient of  $.50$ .

For example, let us say that our hypothetical researcher found that the correlation coefficient between sleep and test performance was  $.80$  for math and  $.73$  for spelling. Such a finding indicates a strong relationship between these two achievement test score variables. With this strong a relationship for both math and spelling, knowing the level of reported sleep allows us to speculate what test performance is. If a correlation coefficient of  $.45$  was obtained for the math test score only, it indicates a moderate relationship. We would be less capable of predicting a student's test score on math knowing his or her level of sleep per night.

You might be surprised to learn that a correlation coefficient of  $-1.0$  translates to a model of perfect prediction, since it involves a negative correlation. A correlation coefficient of  $-1.0$  indicates a perfect inverse relationship. An inverse relationship (also called a negative direction) means that high scores on one variable result in lower scores on the other (i.e., the inverse of a positive direction). Suppose the researcher had obtained a strong negative correlation between sleep and test scores. The prediction of test scores given sleep patterns would be the inverse of what was stipulated above for a positive correlation. If a correlation coefficient of  $-.80$  was obtained for math and a coefficient of  $-.90$  for spelling, the following relationship can be detected for both test scores: The more

sleep a student gets, the lower the test performance. Conversely, the less sleep he or she gets, the higher the test scores. To reiterate, with negative correlations of sufficient magnitude, high scores on one variable predict lower scores on the other variable.

In summary, a correlation coefficient is the degree to which the collected data fit into a linear equation, that is, a straight line. A more useful statistic is produced when you square the correlation coefficient (i.e., you multiply it by itself). The resulting statistic is called  $r^2$ . This statistic,  $r^2$ , depicts the percentage of common or shared variance between the two variables. For example, a correlation coefficient of .90 depicts a strong positive correlation between two variables. To obtain the percentage of common variance, you square it, multiplying .9 by .9, thus getting .81 or 81%. Using this example, the obtained relationship depicts a large amount of shared variance between these two variables. It means that 81% of the variance in the data sets is accounted for by the relationship between reported levels of sleep and test scores. Conversely, 19% of the variation is unaccounted for, meaning the scores are accounted for by some unknown (unmeasured) predictor variable.

Table 1.6 presents correlation coefficients that are hierarchically arranged from strong to weak relationships between two variables, providing the percentage of common variance in the second column.

Table 1.6

### EXAMPLES OF CORRELATION COEFFICIENTS AND SHARED VARIANCE

CORRELATION COEFFICIENT ( $r$ )	COMMON OR SHARED VARIANCE ( $r^2$ )	
.95	$.95^2 = 90\%$	Strong relationship—much common variance
-.92	$-.92^2 = 85\%$	Note that negative correlations still indicate strong relationships between two variables
.77	$.77^2 = 59\%$	Note the drop in common variance when the correlation coefficient is lower
-.31	$-.31^2 = 10\%$	Little common variance
.20	$-.20^2 = 4\%$	Negligible common variance
.09	$-.09^2 =$ less than 1%	Virtually no relationship; one score cannot provide you with any idea about a score on another variable

A rough framework for assessing the strength of a relationship between two variables on the basis of the correlation coefficient (irrespective of sign) is as follows: (1) correlations between 0.00 and .35 are low, (2) correlations between .35 and .65 are moderate, and (3) correlations above .65 are strong (Gay et al., 2006).

Many correlation studies use already existing (archival) data on two or more variables. This merely requires the researcher to compile the data sets across the participants of the study and then subject this to statistical analysis. For example, studying the effects of early intervention on the social development of kindergarten students with developmental disabilities would attempt to find out if a relationship exists between receiving early intervention and social development. The researcher would identify two groups of kindergartners with disabilities, those who received early intervention and those who did not receive it. Some measure of social development would already be in existence. This data would be regressed against the variable of early intervention (dichotomous variable). These types of studies have been referred to as causal-comparative studies, although the term causal is a misnomer. One cannot infer cause and effect from any study whose methodology did not experimentally manipulate the independent variable.

## Conducting a Correlation Study

Conducting correlation research is relatively straightforward in terms of the data collection process (see Table 1.7). Very often, such studies use data that is already existent (called archival data). Selection of the variables that you believe are related and the pool of participants from which to extract the data set is an early decision in the research process.

## Utility of Correlation Studies for Classroom Practice

Prediction studies can be useful where there is a high correlation between predictor variables and some criterion you want to predict. Formulas that determine a person's creditworthiness are used to make judgments about loans and the amount of the loans. It would be silly to simply "go with your gut" when dealing with hundreds of thousands of dollars in prime loans for residential or commercial buildings. Correlation coefficients and prediction models allow the user to possibly predict (with some minimal level of error) what score you would get on one variable by knowing another. Using such predictor models allows people in decision-making positions

Table 1.7

**STEPS TO CONDUCTING A CORRELATION STUDY**

- Delineate the two variables you believe are associated with the population of interest (one of the variables being relevant for student learning behavior).
- Define the group of students who will be included in the data collection process.
- Measure the student variable in a quantitative fashion across many students from the designated sample representing the population of interest; if the measurement has already been performed, extract the data from archives.
- Define the other variable (instructional approach, teacher behaviors or characteristics, school characteristics, and so forth).
- Measure, in a quantitative fashion, that variable or those variables if it is a multivariate study.
- Subject the set of data to a statistical test that determines the degree to which the student variable is related to the teacher or school variable, called a correlation coefficient.
- Determine whether the obtained coefficient indicates a strong relationship by determining if the result is statistically significant (e.g., at the .05 alpha level of significance).

to make fewer errors in judgment. In another example, college admissions departments at major universities try to use certain predictor variables in determining whether a high school graduate would be successful in a particular college program.

However, there is a major caveat to correlation research and its utility for classroom practice. No matter how high the correlation is between the two variables, one cannot conclude that a cause-and-effect relationship exists. Correlation research does not allow for an analysis of cause and effect. For example, let us say that the hypothetical study referred to above finds that sleep level is positively related to school performance. In other words, the larger the amount of reported sleep, the higher the test scores of students with ADHD. Does that mean this researcher proved that getting more sleep will make someone get better test grades? While it may be tempting to make such a claim, this conclusion goes beyond the research methodology and resulting data. It requires a leap of faith. An alternative causal explanation could be that better school performance may cause better sleep. Or possibly another variable, number of hours spent studying, causes both better performance and more hours of sleep. With a correlation study, one does not

know what causes what. Correlation does not allow for a determination of causation.

Here is another common example of incorrect conclusions from correlation research. Perhaps you have heard about studies that implicate a given part of the brain as responsible for certain types of psychopathology (e.g., ADHD). Such research has often demonstrated an association at best. Finding that certain areas of the brain show lower activity levels in children with ADHD than in control participants (those without ADHD) only shows a relationship. One cannot know for certain either that certain malfunctions in the brain are the cause of ADHD or that these malfunctions are the result of the child engaging in certain impulsive behaviors over a long period of time. Which came first? Despite the position of some people regarding the uncovering of a causal model of ADHD, only an experimental analysis can address questions of causality.

The mechanism that is lacking in correlation studies and that precludes causal analysis is the failure to actively manipulate the independent variable. Correlation studies involve only the collection of quantitative data on two sets of variables (or more variables in complex studies). No experimental manipulation takes place! Therefore, one cannot discern whether reported changes in the levels of an independent variable cause changes in the scores of the dependent variable. This inability to determine cause and effect exists irrespective of how strong the relationship is (that is, even with a high correlation coefficient). The hypothetical study described above depicting the strong correlation between sleep and test scores would have to be altered to allow for a cause-and-effect analysis. The researcher would have to systematically vary the levels of sleep in the groups of participants and assess its impact on test grades. If the level of sleep is actively manipulated, the examined relationship becomes one of cause and effect. This process of experimental manipulation will be detailed later.

Why are correlation studies not useful in applied classroom research? I stipulate several reasons. First, correlation research studies examining variables that are out of the province of the teacher are of little use to him or her. Demonstrating a relationship between parental involvement and student achievement probably will not help those teachers whose parents do not show up to parent-teacher conferences. They cannot trade in their set of parents for another set of parents, ones who are more involved with their child's education. Further, a relationship between parental involvement and student achievement,

even if a strong one, does not preclude the possibility that maximizing some instructional or environmental classroom variable can override the influence of a student's home involvement in students who come from uninvolved parents. The proliferation of correlation research with variables that have little utility in classroom practice has probably contributed to practitioners' attitudes about the lack of utility of most educational research.

Correlation studies are irrelevant for practitioners for another reason. Practitioners in classrooms and schools are faced with the task of changing significant social and academic behaviors of the students they serve. Theory building is not a requisite for the classroom teacher. But getting students who are behind established grade level standards in reading to progress to grade level standards is of relevance. For teachers, research that shows a cause-and-effect relationship between an instructional method and student learning is more relevant to practice. Many people espouse the idea that contingent praise for appropriate behavior among elementary-age students is a great technique to get them to stay on task and engage in appropriate classroom behavior. Fortunately, it is more than a wishful theory. Multiple studies conducted in the 1960s and 1970s demonstrated that when the teacher was taught how to provide contingent praise for attending there was a causal effect on student behavior. The pragmatic motivation of personnel in the field would appear to depend on the utility of experimental research.

## EXPERIMENTAL RESEARCH STUDIES

### Questions to Ponder

- What is a functional relationship?
- How would you demonstrate a functional relationship between length of weekly practice time and percentage correct on tests of memorization of multiplication facts?
- Why does a researcher need to randomly assign participants to groups in group designs? What might happen if one group was markedly different on an important variable from another group in a treatment study?
- What are confounding variables?
- What is random selection? How is it different from random assignment? What does random selection allow for? What does random assignment control for?

- What are some problems with random selection of participants in applied settings?
- What are some of the limitations of group designs that you feel make them not feasible for use in school settings?

## Demonstration of a Functional Relationship

To really understand student behavior is to be able to say that A causes B, not merely that B follows A (Bailey & Bostow, 1981; Cooper et al., 2007). Demonstrating that something causes something requires that the researcher actively manipulate one variable and determine whether changes in the other variable occur. This is termed a *functional relationship* (Cooper et al., 2007). A functional relationship is demonstrated when active changes in one variable (the independent variable) produce concomitant changes in a dependent variable. Demonstrating a functional relationship between the independent variable and a dependent variable is the gold standard for discerning treatment and intervention effects. It allows you to reach conclusions on cause and effect by manipulating the independent variable and studying how such a change affects the dependent variable.

To contrast experimental research methodology with a correlation study, let's again examine the effects of sleep on test performance. The study previously referred to demonstrated a relationship between these two variables. However, we cannot conclude that getting ADHD students who have poor test scores to get more sleep would result in increased test scores. To address that question, the researcher must actively alter the level of sleep (must have at least two levels of this variable) and see if this produces differences in the test scores.

The hypothetical researcher begins her scientific inquiry by gathering a specific sample of ADHD students. To determine whether increasing sleep will affect test performance, she selects only those students with poor test grades who concurrently slept less than other ADHD students. In order to actively manipulate the amount of sleep and study, she divides this sample into two groups. The assignment of any student from the initial sample to one of the two groups is random but results in an equal number of students in each group. The study has two phases. One group (called the *control group*) receives no special instruction across both phases of the study. The members of this group continue recording the amount of sleep they get, and their teacher records their test grades at the end of the study.

The other group is the *experimental group*. For the first half of the study, the members of this group receive no special instruction, probably continuing their same pattern of sleep. But for the second phase of the study, they are given a number of strategies to get more sleep. Hence, during the second phase of the study, Group A, the control group, continues with low levels of sleep per night, whereas the members of Group B, the experimental group, after receiving the intervention procedures, increase their sleep by an average of 2.5 hours. This methodology allows for a within-group comparison of test grades across the two phases of the study for the experimental group.

However, the big question is this: Does getting more sleep produce better test grades? The members of Group B may demonstrate this if their test grades improve from the first phase of the study to the second phase of the study. Group A serves as the control, for comparison purposes. Its members' scores should not change substantially if the intervention is solely responsible for increasing test scores. One can therefore demonstrate that changing sleep duration does (or does not) affect test performance. The percentage gain across all participants in each group would answer the question empirically. Table 1.8 outlines the research process in this fictitious experimental study and presents hypothetical data on the mean percentage gain for the control and experimental groups.

The control group does not demonstrate much of a change between the two phases, I and II, in terms of test performance (see right-hand column in Table 1.8). In contrast, the experimental group shows a substantial change in mean test performance, as reflected in the right-hand column. If the difference between 4% and 25% is statistically significant, we can say that getting a student with ADHD (who is not getting enough sleep) to sleep more causes an increase in that student's test grades. As this example illustrates, cause-and-effect studies involve an active manipulation of the

Table 1.8

**RESEARCH PROCESS**

	PHASE I	PHASE II	MEAN % GAIN
Group A (control group)	Baseline condition	Baseline condition	4%
Group B (experimental group)	Baseline condition	Treatment condition	25%

independent variable, also called the treatment variable. The researcher then determines whether such a change produces a change in the data obtained from measuring the dependent variable.

Here is another example demonstrating experimental manipulation of the independent variable. A teacher of American history often puts on costumes of the period she lectures about. She calls this teaching through dramatic reenactment. She has seen it on several TV shows and the students (actors) on TV all respond positively. She vehemently asserts that this is an effective technique for student learning. Does acting out periods of American history in dramatic fashion improve students' retention and comprehension of historical facts? While it may look good, does it really result in improvements in student learning? Only an experimental research study would answer that question. Such a study would determine the effects of dramatic reenactment (independent variable) on student retention and comprehension of historical facts (dependent variable).

The researcher will have to construct several tests of American history content across several weeks of lesson plans. She selects an even number of classrooms to participate in the study. In half of the classes, the teachers receive in-service training from her on using drama and costumes in their history classes. They are provided with suggestions for dramatic reenactment for each of their lesson plans. The other half of the classes do not receive any training. The comparison will be between these two groups of teachers, in terms of the students' scores on the American history tests. If the group using dramatic reenactment has history test scores that are significantly higher than those of the other group, then one can conclude that reenactment caused the change in test scores. If there is no significant difference, it did not.

Experimental research is essential to the development of an effective technology of learning and behavior change. *Too often, suggestions about effective teaching or instructional procedures are based on correlation studies.* Instructional procedures should be subjected to experimental analysis before they are counted as studies providing an evidence base for teachers. Teachers need to know that a treatment, intervention, or strategy is effective, that is, changes student behavior and/or improves student learning. Research that is only correlational in nature leaves too much of a leap of faith in terms of efficacy and does not build an evidence-based approach.

There are two forms of experimental research designs: (1) group designs (between-participant effects) and (2) single-case designs (within-subject effects). The remainder of this chapter will primarily focus on

group designs, since single-case designs are the focus of the rest of the book.

## Between-Participant Effects: Group Designs

Group designs are experimental designs that determine between-participant effects. The effects of the independent variable are determined by comparing the scores of different participants on the dependent variable. The basic methodology is to randomly assign participants (from a pool of participants) to different groups. Differences between two or more treatments are determined on the basis of different test scores or student behavior. Table 1.9 provides three hypothetical examples of such studies with multiple levels of an independent variable. The independent variable can be called the treatment variable. The dependent variable will be specified as a quantitative measure of student disruptive behavior.

In the first row of Table 1.9, labeled A, the proposed study will examine the differential effectiveness of two different strategies for dealing with disruptive student behavior. A strategy termed referral to the principal for behavior problems will be compared with a time-out contingency for disruptive behavior. The level of disruptive behavior for the students in the group receiving the referral intervention will be compared with the level of disruptive behavior for the students receiving the time-out intervention.

Other comparisons can also be made. For example, in a different study, three treatments will be compared (see Table 1.9, row B). In this

Table 1.9

### THREE RESEARCH STUDIES EXAMINING VARIOUS INTERVENTIONS DEALING WITH DISRUPTIVE STUDENT BEHAVIOR

ROW	RESEARCH STUDY	LEVELS OF INDEPENDENT VARIABLE	EXPERIMENTAL CONDITIONS
A	Study of referral vs. time out	2	Referral vs. time out
B	Study of referral vs. time out vs. ignoring	3	Referral vs. time out vs. ignoring
C	Study of referral vs. weekly counseling sessions	2	Referral for disruptive behavior vs. weekly counseling sessions

proposed study, the effectiveness of referral versus time out versus ignoring the behavior will be evaluated according to the rate of disruptive behavior. Therefore, in this study, there are three levels of the treatment variable. In row C of the table, a comparison of two different referral processes, referral for disruptive behavior versus referral for counseling, will be made. The treatment variable (i.e., the independent variable) allocates the number of groups that will be needed to evaluate the number of different treatments (and controls).

Here is another study with a different dependent variable. A researcher posits that pausing after a behavior problem can be an effective strategy for handling student behavior problems. For the purposes of his research, unauthorized talking is the dependent variable. In Table 1.10, four possible studies are delineated. In the first row, two treatments are contrasted. One group will receive the pause intervention while the other will not. In row B, the researcher will have three groups. One group will receive the pause intervention. Another group will receive an intervention termed *reprimand for unauthorized talking*, while the third group will receive a response cost treatment.

In the fourth row (D), four different treatments will be compared (requiring four different groups). This study will compare different rates of pauses for each displayed target behavior. One group will get a continuous schedule of pauses, that is, 1:1, another group will get a pause every other incident of the target behavior, another group will get one pause every five target behaviors, and the final group will get no pauses (control group). This study will evaluate the efficacy of pauses as a function of schedule.

Table 1.10

**FOUR RESEARCH STUDIES**

ROW	LEVELS OF INDEPENDENT VARIABLE	EXPERIMENTAL CONDITIONS
A	2	Pause vs. no pause
B	3	Pause vs. reprimand vs. response cost
C	2	Pause vs. behavioral contracts
D	4	Pause 1:1; pause 1:2; pause 1:5; no pause

Suppose the researcher is interested in knowing whether the intervention is equally effective for male students as for female students. The study becomes more complicated if the effects of the actively manipulated variable are examined on an assigned variable such as gender. We then have two independent variables: treatment and gender. The results will be reported for each gender across the levels of the treatment variable (see Table 1.11). It is also possible that two studies could be conducted, one with male students (Study 1) and the other with female students (Study 2).

## Methodology of Group Designs

In using an experimental group design, several methodological requirements must be in place for the results to have a level of credibility. To have credibility, the results need to be both internally valid and externally valid. Internal validity refers to the methodology being capable of allowing the results (i.e., with regard to the dependent variable) to be attributable to the alteration of the independent variable. External validity refers to the results being generalizable to the intended population, not simply the sample.

### *Random Selection and Assignment*

Those selected for participation in group designs should be randomly selected from the population about which you wish to draw inferences. Random selection from a designated population allows the researcher to

Table 1.11

#### EFFECTS OF PAUSING ON RATE OF UNAUTHORIZED TALKING

##### STUDY 1: MALE STUDENTS IN ELEMENTARY SCHOOL

###### LEVELS OF INDEPENDENT VARIABLE

###### EXPERIMENTAL CONDITIONS

2

Pause vs. no pause

##### STUDY 2: FEMALE STUDENTS IN ELEMENTARY SCHOOL

###### LEVELS OF INDEPENDENT VARIABLE

###### EXPERIMENTAL CONDITIONS

2

Pause vs. no pause

achieve a level of external validity, that is, the results of the study can be generalized to the intended population. Random selection requires that the researcher select those for participation in the study at random from among the intended population.

How does a researcher randomly select students, when all research protocols require the researcher to obtain informed consent from the parents of each child in the study? Random selection is usually not practicable in educational settings and therefore is difficult to achieve for many educational researchers. Most researchers are restricted in some manner in selecting the participant pool for a given study. They do not achieve random selection. Therefore, most experimental studies using group designs have to preface their conclusions with some discussion about the external validity<sup>1</sup> of the findings.

Random assignment is often easier to achieve than random selection, but it controls for a different aspect of the research methodology. In random assignment, participants are selected and then randomly assigned to the different treatment conditions. Random assignment is used to control for potential differences between the participants in each group before the study commences. You do not want differences on relevant variables between groups prior to the different treatments being implemented. For example, if you were conducting a study in which skill aptitude could have a confounding effect, randomly assigning 60 participants to one of the two treatment groups could take care of it. Just like a coin flip, about 50% of the time heads will appear and 50% of the time tails will appear. A table of random numbers is often used, or some other method of random assignment.

What can happen when a researcher does not randomly assign participants to experimental conditions? Let us say you want to determine if vitamin C results in greater strength and endurance. You devise some tests of strength and endurance. You have one group take a placebo pill (not vitamin C but looking like it). You have another group take vitamin C for 12 weeks. You find a difference between the two groups' scores on tests of strength and endurance after the 12 weeks of the experiment. You want to conclude that vitamin C is great for gaining strength and endurance.

But were the groups equal in strength and endurance prior to the study? Could the results be ascribed to a biased assignment of participants to conditions? Suppose the placebo group was culled from four sections of an introductory psychology class, and the vitamin C group was composed of physical education majors in an advanced class for bodybuilders

at the same university. As you can imagine, there are probably vast initial differences between the two groups in strength and endurance prior to the experiment. The probable pretreatment differences in strength and endurance between the two groups create an ambiguous interpretation of the results of this study.

### *Control for Extraneous (or Confounding) Variables*

In experimental group design research, the function of the specific experimental design and method utilized is to provide a level of credibility to the conclusion. Achieving good internal validity allows the researcher to conclude that the independent variable produced changes in the dependent variable. However, alternate conclusions may be possible if the researcher did not control for possible extraneous variables or confounds. What is an extraneous or confounding variable? Think of confounding variables as alternate possible explanations of the obtained results. As an example, suppose an educational researcher assesses the math performance of students in 10 elementary school classes for a 5-week period. Her dependent variable, math performance, is measured by the students' scores on a summative math test that covers the previous 5 weeks' content. She has a theory about the students not doing well in math class. She believes that their blood is not circulating enough through the extremities, which affects the brain flow as well. Her conjecture is that some brief physical exercise prior to math period would get their blood flow going, producing more blood flow to the brain. The result should be better performance.

She initiates the study by collecting the test scores of each student for the first 5-week condition (Condition A test scores). She then provides 10 minutes of exercise for the students in all her classes before the math period for the next 5 weeks. The students take a second math test (producing Condition B test scores), this time on the content of the 5 weeks in which exercise was provided regularly. She finds that the students performed better on the second math test, that is, the test administered after the exercise condition. The average gain in test performance was 10%. She is elated that her theory has seemingly been proved accurate, and reports to the scientific community in a paper presentation that physical exercise before math results in higher math scores. Can one conclude that providing exercise causes math performance to increase? Or are there alternative explanations?

During the researcher's paper presentation, someone in the audience asks if the math period was scheduled during the morning, and if so, at what time. The teacher reports that prior to the Condition A test scores, the math period was after lunch. It remained that way for the first week of the second condition (physical activity). However, due to scheduling changes, the math period was moved to the first instructional period of the day for the remaining 4 weeks of the exercise program. A sigh goes up from the audience. Could it be that higher test scores were achieved because the second phase of the study provided math instruction in the morning, as opposed to the afternoon? As you can see, when another variable, such as time of math instruction, varies systematically with the independent variable, it creates ambiguity with respect to determining cause of change in the dependent variable.

Are there other explanations for why the test scores improved from Condition A to Condition B? Could the content have been easier during Test B? If that was the case, would a better interpretation of the change in test scores be that the change was more a function of the change in content? As you can see, a confound exists in regard to an unambiguous interpretation of the independent variable effects on the dependent variable. Alternate explanations provide a possible explanation of the changes in the test scores of the students, that is, the dependent variable. There are several types of extraneous variables that educational researchers have to control or account for when using group designs (see Table 1.12).

## Conducting a Between-Groups Design Research Study

The steps in conducting a group design cause-and-effect research study are presented in Table 1.13.

Table 1.12

### EXTRANEOUS VARIABLES

- Could the results be a function of participant maturation?
- Could the results be accounted for by sensitization to test conditions (when pre- and posttest used, called a practice effect)?
- Could the results be ascribed to biased assignment of participants to conditions?
- Could the results be accounted for by increased loss of participants in one condition versus another (called experimental mortality)?
- Could the results be due to participant expectancy effects?

Table 1.13

**STEPS TO CONDUCTING A GROUP DESIGN STUDY**

- Delineate the dependent variable for the study (a variable that is relevant for student learning or behavior).
- Define the group of students who will be used in the data collection process, with the dependent variable being some measure of their behavior or skill acquisition.
- Define the independent variable (instructional approach, teacher behaviors or characteristics, school characteristics, and so forth).
- For any given independent variable, identify the treatment(s) and control condition that will be tested in your study.
- Determine which students will be assigned to which treatment or control group.
- Implement the treatment phase of the study.
- Measure the student variable in a quantitative fashion across all the students in their assigned group (the time such data is collected is determined by the actual design used).
- Subject the set of data from the dependent variable to a statistical test that determines whether the difference between the treatment group(s) and the control group is statistically significant (this is called the alpha level, usually set at .05).

**Utility of Group Designs for Classroom Teachers**

While group designs provide an experimental methodology allowing a researcher to conclude that a treatment variable produces changes in the dependent variable, they are fraught with logistical problems. There are five practical limitations of group designs that make the use of such a research methodology impracticable for classroom personnel. The limitations of group designs appear to have encouraged the proliferation of single-case designs over group designs in applied settings.

The five limitations of group designs that appear to restrict their use in many applied settings and classrooms are presented in Table 1.14.

The first practical limitation of a group design is an ethical objection to the withholding of any effective treatment to participants in a control or baseline condition. A control or baseline condition involves the giving of no treatment to those groups of participants (withholding of treatment). The control group is used to evaluate an experimental treatment against possible nonspecific treatment effects and can be considered to

Table 1.14

**FIVE LIMITATIONS OF GROUP DESIGNS**

- Ethicality of withholding treatment from control participant
- Impracticality of collecting a large pool of participants with same target behavior problem
- Impracticality of obtaining random samples
- Unavailability of data comparing rate of change within each participant
- Unavailability of magnitude of treatment effects for individual participants

function in a manner quite similar to the baseline condition in single-case designs. Therefore, participants assigned to the control condition in a group design do not receive treatment. In contrast, single-case designs allow for all participants to be exposed to each experimental condition and (hopefully) ultimately experience a desired change of behavior.

As an example of the first limitation of group designs, let us consider a two-group design that would experimentally examine the effect of phonics lessons on spelling tests. Such a study could be constructed in the following manner: One group would receive 20 minutes of phonics instruction. Another group would not receive such a treatment (they might receive an alternate instructional activity). As you can see, an ethical dilemma presents itself. All students in the class should be entitled to a procedure, that is, phonics, that can improve their spelling. This can be partially attended to by providing treatment after the data is collected, but some students would receive a benefit only for a shorter period of time than others.

A second limitation is the impracticability of obtaining the large number of participants necessary for group designs. If teachers are going to conduct research in their classrooms, they are hampered by the number of students in their class. If you have 20 students and want to compare four treatments, you would only have 5 students in each group. This is a small number with which to produce a statistically significant effect between one or more conditions.

A third limitation for the classroom teacher is the impracticability of obtaining random samples of participants. Group designs require random selection for the generality (generalization) of results. If this is not obtained (and it is usually difficult to obtain in an applied setting), the generality of the findings is limited. In evaluating treatments

in classrooms, where the needs of individual students are of concern, it becomes unfeasible to obtain random samples for methodological purposes. Random selection is difficult for a single teacher to conduct, since the students in his or her class are not randomly selected from a target population.

A fourth limitation is the use of a single score or measure for each research participant. This does not allow for a comparison of the change that might have resulted from a change in experimental condition with any given participant. In order to individually determine how much change a given experimental condition produces, it is essential to know what the rate of the behavior was before the experimental condition was implemented, and the rate of the behavior as a result of the experimental condition. Group designs that collect only one score for each individual do not allow for a before- and after-treatment comparison for each participant.

The inability of group designs to allow for individual analyses of experimental effects (Bailey & Bostow, 1981; Cooper et al., 2007) leads to a fifth limitation. Group design methodology determines effects by averaging the scores of a number of participants. This does not allow for examination of intersubject variability of treatment effect. It is conceivable that the treatment procedure may affect some participants considerably, whereas others may be less affected (or not at all). Single-case designs, due to their measuring of the effects (behavior change) as a function of a number of different conditions, can provide data on the size of the experimental effect for each participant.

Due to these limitations, it is advisable that teachers conducting research in their own classrooms utilize single-case designs to minimize the amount of time and cost in evaluating a number of alternative treatment conditions. Group design does have a place in education research, but this methodology seems to be more appropriate for large-scale efforts to crunch achievement test data across many schools or districts.

## **Introduction to Single-Case Designs**

Single-case designs, like group designs, actively manipulate the treatment variable. The researcher can discern whether a given treatment produces a change in student behavior and/or learning. Single-case design is the generic term used to refer to a cause-and-effect research paradigm that has three distinctive characteristics (Bailey & Bostow, 1981; Cooper et al., 2007): (1) repeated measurements of the same behavior within a

single experimental condition; (2) implementation of each treatment or experimental condition across each participant; and (3) use of the individual participant as the basis for comparison of experimental effects. In contrast to what appears to be implied by the term, single-case research designs may utilize more than one participant. You can conduct a single-case research study with many students in your class, provided the three distinctive characteristics are evidenced.

The first characteristic involves collecting data repeatedly on the same behavior (dependent variable) over a period of time. For instance, if one were to measure a student's noncompliant behavior to teacher directives in the classroom, the measurement of this child's behavior would occur repeatedly within the same condition. The child would be observed in a number of different sessions or on a number of different days, rather than being observed just once. The same behavior, noncompliance with teacher directives, is measured across time, rather than the measurement of several different behaviors being alternated. The repeated measurements within a given experimental condition can be made during multiple sessions, days, weeks, or even blocks of trials. However, each new session, day, week, or block of trials (whatever is being used) should represent another data point on a graph.

The second characteristic of a single-case design is that multiple experimental conditions (all having repeated measurements) are implemented across each participant. As was pointed out earlier, single-case designs can utilize more than one participant in the research study. Further, the study can compare different experimental conditions. However, each participant is exposed to more than one experimental condition. This allows for a comparison of any two conditions on the individual student's behavior. The determination of effectiveness is made for each individual participant.

Single-case designs measure within-participant experimental effects, in contrast to the study of between-participant effects in the case of group designs. Why is it important to assess within-participant effects? Some treatments may be effective for some students but not for others. If you average the results across an entire group, individual differences in treatment effectiveness are lost. In contrast, if each individual participant serves as the basis of comparison for each level of the independent variable, individual variation with respect to each condition will become apparent.

The remaining chapters in the book deal with single-case designs. Chapter 2 covers measurement methods for observational research, which is the primary method of measurement in single-case designs.

Chapter 3 details the characteristics of single-case designs, and it is followed by a chapter on single-case experimental designs. Chapter 5 presents information on possible categories of single-case research studies.

## NOTE

1. This is why direct replication over many studies is needed in verifying treatment efficacy.

## REFERENCES

- Bailey, J. S., & Bostow, D. E. (1981). *Research methods in applied behavior analysis*. Tallahassee, FL: Copy Grafix.
- Cohen, L., Manion, L., & Morrison, K. (2003). *Research methods in education* (4th ed.). New York: Taylor & Francis Routledge.
- Cooper, J. O., Heron, T. F., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Columbus, OH: Merrill/Prentice Hall.
- Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational research: Competencies for analysis and applications* (8th ed.). Columbus, OH: Merrill/Prentice Hall.
- Gunter, P. L., Reffel, J. M., Barnett, C. A., Lee, J. M., & Patrick, J. (2004). Academic response rates in elementary-school classrooms. *Education and Treatment of Children*, 27, 105–113.
- Matheson, D. W., Bruce, R. L., & Beauchamp, K. L. (1978). *Experimental psychology: Research design and analysis* (3rd ed.). Fort Worth, TX: Holt, Rinehart & Winston.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.