SECOND EDITION

METHODS IN EDUCATIONAL RESEARCH

From Theory to Practice

MARGUERITE G. LODICO • DEAN T. SPAULDING • KATHERINE H. VOEGTLE

Methods in Educational Research

Methods in Educational Research

From Theory to Practice

SECOND EDITION

MARGUERITE G. LODICO
DEAN T. SPAULDING
KATHERINE H. VOEGTLE

JOSSEY-BASS
A Wiley Imprint
www.josseybass.com

Published by Jossey-Bass
A Wiley Imprint
989 Market Street, San Francisco, CA 94103-1741—www.josseybass.com

Library of Congress Cataloging-in-Publication Data

Lodico, Marguerite G.
Methods in educational research : from theory to practice/Marguerite G. Lodico, Dean T. Spaulding, Katherine H. Voegtle.—2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-43680-6 (pbk.)
1. Education—Research—Methodology. I. Spaulding, Dean T. II. Voegtle, Katherine H. III. Title.
LB1028.L586 2010
370.72—dc22
2009051930

Printed in the United States of America
SECOND EDITION
PB Printing 10 9 8 7 6 5 4 3 2 1

CONTENTS

BOXES, EXHIBITS, FIGURES, AND TABLES

Boxes

Exhibits

Figures

Tables

In loving memory of our colleagues Huey Bogan and Mark Ylvisaker. Huey inspired his students in Teacher Education to be reflective practitioners who continually strive to improve their practice. Mark exemplified the integration of research and clinical practice in his numerous publications and inspired his students in Communication Sciences and Disorders to become scholar practitioners. We miss them both dearly.

PREFACE

Four years ago we wrote the first edition of *Methods in Educational Research: From Theory to Practice* with many expectations. As educational psychologists, we wanted to write a textbook from the fundamental perspective of how one learns in general and, more specifically, how one learns through conducting research. In addition, we wanted to create a book that would pertain to techniques and instructional practices underlying good teaching as well as to teach people about educational research. We wanted to pay close attention to the metacognitive processes associated with learning about research and developing and becoming an active participant in the educational research community.

We believe the purpose of this second edition is still to assist students, primarily graduate students, who are practitioners in education or related fields (administration, school psychology, or school counseling) to use educational research so that they can become more effective educators. Specifically, the purpose of this book is to help students develop a broad and deep understanding of research methodologies that can be used to analyze and improve their practices. Overall, we believe that we accomplished much of what we set out to do with the first edition of this book; however, in using it for the past three years, we realized that some areas needed to be expanded. This second edition expands on areas previously discussed as well as brings you several new chapters. We hope you enjoy the additions and wish you the best in your educational research pursuits.

In chapter 1, we have updated information on No Child Left Behind (NCLB) and school reform as well as the knowledge and research skills needed by educators in the 21st century. Chapter 2 has been expanded to provide more in-depth coverage of key concepts that we believe students need to be able to read and understand the research on a particular topic or issue. Chapter 3 focuses specifically on descriptive statistics and provides students with detailed examples of basic statistical computations and how results are displayed. Chapter 4 has an expanded section on archival data and descriptive statistics has been moved to chapter 3. Chapters 5 through 7 have been expanded, focusing on the different quantitative and qualitative types of research, with specific data exercises embedded in each chapter to give students a realistic framework for the types of data and possible analyses they may use when conducting research. We have embedded sampling strategies for descriptive survey research in chapter 7. In chapter 8, we expanded types of qualitative research to those most relevant to practitioners. Chapters nine through eleven are similar to those in the original edition. Perhaps the biggest change we have made to our book is chapter 12, which focuses entirely on action research. The chapter includes a data activity to give students a sense of the action research process. While the action research chapter is placed late in the book, there are no concepts in it that preclude coverage earlier in the semester. Chapter 13, program evaluation, features a section on logic modeling. Last, chapters 14 and 15 focus on generating ideas and researchable topics and preparing the research proposal. We moved these chapters to the end of the book to give instructors more flexibility in when they are assigned.

The book includes special features designed to assist the teaching and learning processes:

- ❒ Research vignettes illustrating research that is tied to practice and used to make decisions about educational practices open each chapter on research approaches and are discussed throughout these chapters.
- ❒ The book includes extensive discussion of research issues and concepts relevant to the accountability movement and using data to make decisions in educational settings.
- ❒ Developmental processes involved in researching and writing a research proposal are emphasized.
- ❒ Research proposals using both an action and a descriptive survey approach are included in appendices, because we feel these approaches are useful to practitioners. The appendices also include criteria for evaluating proposals using these approaches.
- ❒ Key concepts students should know are set in bold letters in each chapter.

- ❒ Suggested readings are provided at the end of each chapter to extend the discussion of general issues raised in the chapter and provide citations for sample studies that illustrate the type of research discussed.
- ❒ Discussion questions or activities are provided to stimulate thinking about the issues raised in the chapter or encourage students to apply the concepts presented.
- ❒ We did not include sample studies in the book; however, each chapter discussing a specific research approach (for example, descriptive survey, action, experimental) includes a list of studies that may be used for class discussion or assignments. Since many studies are available as full-text documents, we decided to decrease our carbon footprint by encouraging students to access these online.

ACKNOWLEDGMENTS

Any textbook on educational research owes a debt to the numerous people who have built the rich and varied literature in this field. In some sense, this book grew out of the conversations and relationships we have enjoyed with colleagues over many years, especially at meetings of the American Educational Research Association and the American Evaluation Association. Although we cannot name all of these persons, we certainly could not have begun to think about this book without the stimulation of many people in these vibrant educational communities.

However, many people closer to home also made this book possible. The College of Saint Rose and especially our dean, Margaret M. Kirwin, provided substantial support for our work by granting us sabbatical leaves and making available to us capable graduate assistants. The revisions of the book were informed by feedback from faculty members who were using it and from students in our classes. Members of our department consistently encouraged us in our writing, and our department chair, Richard Brody, always managed to get people to cover courses as needed each semester. Our colleagues who are practicing educational researchers—James Allen, Aviva Bower, Donna Burns, David DeBonis, Ron Dugan, Margaret McLane, Heta-Maria Miller, Travis Plowman, and Ismael Ramos—each contributed his or her own special expertise and pedagogical ideas to the book.

Moira DeSanta, our graduate assistant, carefully read and edited chapters from our book. We are also certainly grateful to the staff at Jossey-Bass, including Kelsey McGee, Andy Pasternack, and Seth Schwartz, for their continued support.

We also thank the students from our educational research classes who patiently read through often-imperfect drafts of the book, providing feedback and suggestions. Many of them allowed us to include samples of their work in this book to help us fulfill our goal of making courses on educational research more comprehensible, relevant, and useful to future generations of preservice educators. In particular, we thank Robert Dexter and Megan Rempe, whose research proposals are included as Appendix A and Appendix B in our book.

Finally, on a personal level, we thank our partners, Phil Lodico, Evan Seiden, and Jim Fahey, who kept us sane, well fed, and entertained throughout the often hectic job of revising this book.

M.G.L., D.T.S., and K.H.V.

THE AUTHORS

Marguerite G. Lodico received her EdD from the University of Houston and is a professor at the College of Saint Rose, where she teaches child development and educational research. In addition to her teaching responsibilities, she has served as chair of the Educational and School Psychology Department, as interim dean and as director of an after-school mentoring program. She has conducted research on students in urban environments and school-to-college collaborations. She is currently involved in a professional development school initiative with a local elementary school. She is the coauthor of the case book *Child and Adolescent Life Stories: Perspectives from Youth, Parents, and Teachers* published in 2004.

Dean T. Spaulding received his PhD from State University of New York at Albany in 2001 and is currently an associate professor at the College of Saint Rose where he teaches educational research and program evaluation. His research and program evaluation work has focused on technology, after-school and enrichment programs, educational leadership, and environmental and science education in K–12 settings and in higher education. He is the author of *Program Evaluation in Practice: Core Concepts and Examples for Discussion and Analysis* published in 2008.

Katherine H. Voegtle received her PhD from the University of Cincinnati and is currently a professor at the College of Saint Rose, where she teaches courses in human development, educational research, and educational psychology.

She is coauthor of *Child and Adolescent Life Stories: Perspectives of Youth, Parents and Teachers* and has conducted qualitative and quantitative research projects on creative language development, school-based ally groups, mentoring programs, and arts-based educational programs. She is currently working on several action research projects that are part of her college's ongoing efforts to build a professional development school.

Methods in Educational Research

CHAPTER ONE

INTRODUCTION TO EDUCATIONAL RESEARCH

CHAPTER OBJECTIVES

- Become familiar with the recent history of the educational accountability movement and describe the role of research in accountability
- Understand the role of action research in improving teaching and learning
- Explain value-added assessment
- Describe key aspects of the No Child Left Behind Act
- Explain the differences between inductive and deductive reasoning
- Articulate the key differences between knowledge-oriented philosophical frameworks for educational research (scientific realism and social constructivism) and action-oriented approaches (advocacy or liberatory and pragmatism) and begin to define your own framework
- Explain the differences among and provide a simple example of quantitative and qualitative methods of data collection and basic and applied educational research
- Understand the essentials of research ethics and how ethics apply to research questions and methodology

EDUCATIONAL ACCOUNTABILITY AND EDUCATIONAL RESEARCH

At the beginning of the 21st century, the educational research community is again responding to the call for increased accountability in our nation's schools. This call for accountability comes from both within and outside the educational community. Educators, parents, students, communities, and politicians are hopeful

BOX 1.1
Educational Reform and the No Child Left Behind Act

In 1965, the Elementary and Secondary Education Act (ESEA) was passed by the U.S. Congress to achieve three major goals. These goals included the desire to improve Scholastic Aptitude Test (SAT) scores, increase academic proficiency, and close the achievement gap that separated students of color and low-income students from White and more affluent students (Nichols & Berliner, 2007). ESEA provided funding to schools (labeled "Title I" schools) with high poverty levels and large numbers of students of color. In 1983, eighteen years after the passage of ESEA, the National Commission of Excellence in Education published a report entitled *A Nation at Risk: The Imperative for Education Reform in America.* Troubling to all, the report stated that ESEA had failed to achieve its goals and that academic proficiency of U.S. students remained low. *A Nation at Risk* called for additional reforms to increase parental and community involvement, improve achievement, enhance the quality of teachers, and close the achievement gap. While *A Nation at Risk* drew attention from educators, parents, and legislators, it resulted in little change or reform. It was not until 1994, under the administration of President Bill Clinton, that serious educational reform came under increased scrutiny. This occurred with another reauthorization of ESEA entitled *Goals 2000,* which focused greater attention on school accountability. As part of this legislation, schools that developed annual testing practices received financial incentives.

Goals 2000 provided a skeletal foundation for the next iteration of ESEA, called the No Child Left Behind Act (NCLB). NCLB was passed by Congress and signed into law by President George W. Bush in 2001. The rationale, in part, was based on the fact that in spite of spending more than $300 billion since 1965 to educate youth from low-income families, only 32% of fourth graders could read at grade level, and most of those who could not read were ethnic minorities (U.S. Department of Education, 2005b). Believing that the money spent was not improving education, NCLB was designed to increase accountability of individual schools and states and ultimately reform education.

that the new accountability will result in increased achievement for America's students.

Accountability and educational reform are by no means new in education (see Box 1.1). The newest accountability legislation, No Child Left Behind (NCLB), holds schools accountable for monitoring and reporting student progress based on test scores. Monies for schools are made available for programs that are scientific and reliable (see Box 1.1, number 5), although the federal government's definition of scientific research is very narrow (Neuman, 2002).

The legislation significantly increased the role of the federal government in education and set into place regulations that reached into nearly all public schools in this country. In short, the legislation requires (U.S. Department of Education, 2005a):

1. *Annual testing.* By the 2005–2006 school year, states were required to test reading and math annually in Grades 3–8. By 2007–2008, states were required to develop tests to measure science achievement at least once in elementary school, middle school, and high school. All tests must be aligned with state standards and be reliable and valid measures. Additionally, a sample of the fourth and eighth grades must participate in the National Assessment of Educational Progress testing program every other year in the content areas of reading and math.
2. *Academic progress.* States are responsible for bringing all students up to a level of proficiency by the 2013–2014 academic school year. Each year, every school must demonstrate **adequate yearly progress** (AYP) toward this goal. If a school fails to meet this goal for two years in a row and receives Title I funding (federal dollars), the state must provide technical assistance and families must be allowed a choice of other public schools (assuming there is available space and that the other schools are making adequate progress). If a school fails to meet the defined level of proficiency for three years in a row, it must offer students supplemental educational services, which could include tutoring.
3. *Report cards.* All states must prepare individual school report cards on all schools. These report cards must be made public and must demonstrate progress in reaching the state standards.
4. *Teaching quality*. Currently, the federal government provides money to states and school districts to improve the quality of their teaching forces. Under the NCLB legislation, the federal government has indicated that it will provide greater flexibility in the spending of that federal money.
5. *Reading First.* NCLB offers competitive grants called Reading First that will help states and school districts set up scientific and reliable research-based reading programs for children in kindergarten through Grade 3. School districts in high-poverty areas will be given priority for these grants.

According to the U.S. Department of Education (2005a), the key characteristics of reliable research are

1. A study that uses the scientific method, which includes a research hypothesis, a treatment group, and a control group
2. A study that can be replicated and generalized
3. A study that meets rigorous standards in design, methods used, and interpretation of the results
4. A study that produces convergent findings, for example, findings are consistent using various approaches

These guidelines have significant implications for the way research is conducted in education. Specifically, the legislation calls for researchers to conduct studies with scientific rigor. According to Neuman (2002), NCLB's definition of scientific rigor is consistent with randomized experimental designs—study designs in which persons are randomly assigned to groups that are treated differently. Randomized studies are one approach for establishing causality but may not be appropriate for all research questions. Nearly everyone agrees that research studies should be rigorous and scientific. However, the narrow definition of *scientific rigor* as randomized experimental studies has the potential for greatly limiting the scope of educational research. Furthermore, according to Davies (2003), "Devoting singular attention to one tool of scientific

NCLB requirements and other accountability measures make knowledge of educational research an essential component of professional preparation for all educators. However, to promote creative, innovative, yet sound solutions to current educational problems, future educators must become knowledgeable about a multitude of research approaches that reach beyond those techniques defined as reliable under the NCLB legislation. It is our hope that this book will enable you to participate in ongoing debates about the status and future of education on both national and local levels. We also hope that you will develop skills and knowledge to take part in a much longer and broader tradition: using scientific research to identify, develop, and assess effective educational practices. Furthermore, by using this knowledge you will be better able to make informed decisions based on data and evidence collected in your practice (for example, what is often referred to as *evidence-based practice*).

It is our belief that practitioners can have a major role in influencing positive change in their classrooms, schools, and districts if they actively engage in the research process. This does not necessitate that practitioners become involved in large-scale research projects. We are all aware that teachers and other educational research jeopardizes inquiry efforts into a range of problems best addressed by other scientific methods" (pp. 4–5).

A school's failure to meet its AYP has serious consequences; these consequences become more severe the longer it takes schools to reach their defined benchmarks. (*Benchmarks* are predetermined levels of achievement for which states or federal officials set performance levels.) For example, a school that fails to make its AYP two years in a row is labeled as a **school in need of improvement** or a SINI school. The SINI school must then develop an improvement plan that describes the necessary changes that will result in meeting its AYP. SINI schools must offer public choice to their students, allowing transfer in-district to a school in good standing or to a nearby charter school (*charter schools* are public schools funded with tax dollars that permit some flexibility regarding some state education regulations). Schools that fail to make AYP for three years must provide and pay for supplemental educational services for eligible students. This often includes tutoring services offered by approved providers. Those SINI schools that continue to fail to meet AYP for four consecutive years must take "corrective action" in addition to the sanctions noted. This action could include replacing administrative staff, hiring outside consultants to run the school, implementing a new curriculum, and extending the school year, to name a few. Schools that fail to make AYP for five consecutive years must develop a restructuring plan that may result in a state takeover or new governance for the school.

professionals have very heavy workloads. In spite of this, many practitioners currently conduct small-scale research projects to evaluate their own practices. This type of research is often referred to as **action research** or practitioner research which is discussed in depth in Chapter 12. Briefly, action research (see Box 1.2 for an example) is a type of research that is conducted by the practitioner in order to improve teaching and learning. Action research is conducted by teachers, counselors, school psychologists, speech language pathologists, administrators, or any educational professionals looking to improve their practice. It is often done in a collaborative environment in which practitioners engage in a cycle of reflection and action to gain knowledge about ways to improve their practices. More specifically, action research provides practitioners with a process that involves reflection or assessment of needs, utilization of a systematic inquiry, collection and analysis of data, and informed decision making.

Action researchers strive to find solutions that can bring immediate change and facilitate improvement in student learning. One might ask why practitioners are increasingly involved in action research. The answer is quite simple. Schools and school districts are involving practitioners to a greater degree in the operation of

BOX 1.2
Action Research Example

Ms. Lovett, a first-year teacher, is teaching a ninth-grade biology unit on parts of the human respiratory system. On the first quiz, which covered the initial part of the unit, 50% of her students failed the quiz. She reviews her quiz and finds it to be fair. Her next step is to reflect on the strategies she used to cover the content. She realizes that her primary instructional strategy was lecture, multimedia, and student note taking. After talking to colleagues and researching best practices, she decides to develop an alternative instructional approach. Her plan of action involves introducing students to a new biology computer software program that allows them to see, through computer animation, the functions of each part of the respiratory system. She decides that for the next section in the respiratory unit, she will take her students to the computer lab. While in the lab, students will spend half the class working with the new software, and for the second half of class she will continue to use lecture, multimedia presentations, and note taking. Ms. Lovett administers a second quiz to the students after two weeks of study. This time only 10% of the students fail. She decides that she will continue to incorporate computer time for the next unit and continue to monitor and assess all the students through the next unit, with a special focus on those who did not improve their performance.

schools. Additionally, practitioners are being held accountable for student learning. These factors have increased the level of participation of the practitioner beyond his or her traditional responsibilities. Practitioners are assessing their own practices and, where appropriate, modifying those practices. Most important, engaging in action research empowers practitioners. They can identify their own practical research problems and set in motion immediate plans to improve practices. This immediacy is attractive to practitioners who are looking to make quick yet responsible and defensible changes or improve the learning of their students. As you read this book, we hope that you will appreciate the importance of considering ways in which the practitioner can use research to make a difference in the quality of our educational systems.

Results of NCLB and New Directions in Accountability

NCLB has both supporters and critics. Those who support the legislation believe strongly that the regulations and accountability through standardized testing will increase student achievement and close the achievement gap (goals of the original ESEA and each of its iterations). However, there are many strong and very vocal critics. Much of the criticism focuses on the use of standardized testing as a single measure of accountability; the results of current research suggest that NCLB is not achieving its goals.

For example, Lee (2006) conducted a comprehensive study and systematic trend analysis of national- and state-level public school achievement in math and reading during the pre-NCLB years (1990–2001) and post-NCLB years (2002–05). The study analyzed achievement across socioeconomic and racial groups with an eye on determining whether the gap in achievement was closing and whether all groups were on target to meet the goals of NCLB (100% proficiency by 2014). Primarily utilizing the National Assessment of Educational Progress (NAEP) test data, the study determined the following:

- NCLB has not improved student achievement in reading. A comparison of NAEP pre- and post-NCLB reading scores was flat, indicating no growth or loss in achievement. While there was a slight increase in math scores immediately following the implementation of NCLB, scores returned quickly to pre-NCLB levels.
- The gap in achievement between racial and socioeconomic groups persists.
- While state assessments in reading and math show some improvement (these trends in many cases began prior to NCLB), these improvements are not demonstrated in the only national test of achievement (NAEP).

The debate over NCLB and its use of yearly standardized assessments has resulted in some educators calling for changes in the way students are assessed and schools are held accountable. According to Doran and Fleischman (2005), "The NCLB approach rests on the assumption that assessment data can provide credible information to gauge how effectively schools and teachers are serving their students" (p. 85). While assessment data may in fact be able to achieve such a goal, a concern under NCLB is the kind of data collected and how adequate yearly progress is calculated and then used as a measure of school effectiveness. AYP is the way that states measure the yearly progress schools are making toward the goal of 100% student proficiency in at least reading/language arts and math. It sets a benchmark or minimum level of proficiency that students must achieve on yearly tests of achievement. It should be noted here that this process puts at a disadvantage a school with a high number of students whose beginning achievement levels are much farther below those of their affluent counterparts. AYP is much easier to meet in a school where only 10% of students are performing below state benchmarks than a school with 90% of its students below state benchmarks for achievement.

Fundamentally, AYP is determined by comparing student academic performance on a single standardized test administered from year to year. For example, let's say that Green Elementary schools' third graders fail to meet the AYP in math in 2007. The method used to determine such failure required that the school compare the performance of the 2006 third graders to the 2007 third graders, two different cohorts. The question asked by educators and administrators is "What information does this comparison tell us about individual student growth and teacher effectiveness when it involves comparing different groups of students?" The answer is that it provides little, if any, information about the progress of individual students. As a result, many in the field of education, and more recently the federal government (see U.S. Department of Education, 2006a & b), are calling for **value-added assessment** systems. Unlike the way student progress is monitored under NCLB, value-added assessment allows educators "to examine and assess their [student] learning trajectories as they progress over time through different classrooms taught by different teachers in different schools and districts" (Amerein-Beardsley, 2008, p. 65). In the value-added model of assessment, teachers and administrators are held accountable through the examination of how much value or improvement they have contributed to an *individual* student's learning. For example, in schools using value-added assessment, the growth of individual students can be tracked across teachers and subjects from year to year. The gains or losses of these individual students are then summed to provide a picture of a school or school districts' progress under the value-added model. Table 1.1 displays the data for school district A and school district B. On this assessment, students' scoring levels 1 and 2 are not meeting learning standards. Students scoring 3 or 4 are meeting or exceeding learning standards. Based on the data provided in Table 1.1, which school district added more value according to a value-added assessment system? If you said, "district A," you are correct. While 80% of the students in district A did not meet learning standards in year 2, the level of growth within this group was greater than in district B.

While there are multiple value-added assessment models currently utilized by a variety of school districts across the country, all models recognize that children come into the educational system with a wide variety of backgrounds and skills. Given this fact, examination of a yearly standardized test score does not accurately identify effective and ineffective teachers and schools. According to Doran and Fleischman (2005, p. 85), "The idea behind value-added modeling is to level the playing field by using statistical procedures that allow direct comparisons between schools and teachers even when those schools are working with quite different populations of students." Value-added assessment measures individual student achievement on a yearly basis and calculates a gain score. The gain score is then used as a more fair assessment of effective schooling. The models utilize

TABLE 1.1 Sample Value-Added Data for Two School Districts

District A		
Criteria	**Year 1 (third grade 2006)**	**Year 2 (third grade 2007)**
Level 1	30%	20%
Level 2	50%	60%
Level 3	10%	15%
Level 4	10%	5%
District B		
Criteria	**Year 1 (third grade 2006)**	**Year 2 (third grade 2007)**
Level 1	7%	6%
Level 2	8%	7%
Level 3	63%	65%
Level 4	22%	22%

complex statistical techniques (well beyond the scope of this book) in order to estimate teacher and curricular effects on students. A major question, however, is whether the gain score obtained from a value-added assessment can be attributed to teaching effectiveness. Is it really possible to determine, even using careful statistical procedures, the relative influence of a wide range of variables (such as socioeconomic status, ongoing after-school reinforcement, preschool attendance) and conclude that the gains are due to a teacher, teaching method, or curricular effects? The answer is not yet clearly known. Research studies are being conducted to determine whether these sophisticated statistical models can separate the differential effects of the many variables that influence student progress.

The value-added assessment system is consistent with the way many action researchers evaluate interventions. Action researchers might ask questions like, "How much has Samuel improved following the use of math manipulatives?" and "How does that compare with how much Louisa improved?"

CONDUCTING EDUCATIONAL RESEARCH

Recent accountability efforts are certainly not the first effort to apply scientific methods to educational practices. Since the beginning of formalized education, research has been used to help improve education and to determine how education works in a wide range of situations.

The Scientific Method

Through scientific research, educators hope to obtain accurate and reliable information about important issues and problems that face the educational community. **Scientific research** as applied to education is defined as the application of systematic methods and techniques that help researchers and practitioners understand and enhance the teaching and learning process.

Much like research in other fields, research in education uses two basic types of reasoning: **inductive reasoning** and **deductive reasoning**. Inductive reasoning is often referred to as a "bottom-up" approach to knowing, in which the researcher uses observations to build an abstraction or to describe a picture of the phenomenon that is being studied. Inductive reasoning usually leads to inductive methods of data collection through which the researcher (1) systematically observes the phenomena under investigation, (2) searches for patterns or themes in the observations, and (3) develops a generalization from the analysis of those themes. The researcher proceeds from specific observations to general statements—a type of discovery approach to knowing. For example, a researcher is interested in determining the nature of the interactions that occur between students with disabilities and regular education students who are educated together in a preschool setting. The researcher spends two days a week for six months observing and interviewing the preschoolers. She specifically focuses on the types of activities these two populations engage in together during the course of the school day. She gathers the notes from her observations and interviews and concludes that the students with disabilities and regular education students play together, eat lunch together, and express positive attitudes toward each other.

In contrast, deductive reasoning uses a "top-down" approach to knowing. Educational researchers use one aspect of deductive reasoning by first making a general statement or prediction and then seeking evidence that would support or disconfirm that statement. This type of research employs what is known as the **hypothetic-deductive method**, which begins by forming a **hypothesis**—a tentative explanation that can be tested by collecting data. For example, one might hypothesize that small classes would result in a greater amount of student learning than large classes. This hypothesis would be based on a **theory** or a knowledge base composed of the results of previous research studies. A theory is a well-developed explanation of how some aspect of the world works using a framework of concepts, principles, and other hypotheses. For example, a humanistic theory of education might emphasize strong teacher-student relationships as part of effective learning. Previous research studies may have shown that such relationships are more common in small classes. Therefore, based on the humanistic theory and these previous studies, the researcher in our example may have hypothesized that small class sizes will result in better student learning based on humanistic theory and previous studies. The next step in the hypothetic-deductive approach is to collect data to see whether the hypothesis is true or should be rejected as false. The researcher might compare student learning in classrooms of 15 or fewer students with those of 25 or more students. If students in the smaller classes show a greater amount of learning, the hypothesis would be supported. If the students in the smaller classes do not show a greater learning, then by deductive reasoning, the hypothesis is shown to be false. To summarize, the researcher (1) began with a theory and a knowledge base and used them to form a hypothesis, (2) collected data, and (3) made a decision based on the data to either accept or reject the hypothesis or prediction.

The inductive and hypothetic-deductive approaches to knowing represent two general routes to knowledge used in educational research. Inductive reasoning is most closely associated with **qualitative research** (see Table 1.2) which collects and summarizes data using primarily narrative or verbal methods: observations, interviews, and document analysis. Qualitative researchers are often said to take inductive approaches to data collection because they formulate hypotheses only after they begin to make observations, interview people, and analyze documents. These hypotheses are examined and modified by further data collection rather than being accepted or rejected outright. Qualitative researchers believe that full understanding of phenomena is dependent on the context; they use theories primarily after data collection to help them interpret the patterns observed. Ultimately, qualitative researchers attempt to make claims about the truth of a set of hypotheses, although they may confirm these hypotheses primarily for a given setting or context.

The hypothetic-deductive method is most closely associated with **quantitative research**, which summarizes data using numbers. Hypotheses and methods of data collection in quantitative research (see Table 1.2) are created before the research begins. Hypotheses or theories are then tested and when supported add evidence supporting the theory. Over time, supportive findings with different groups in different settings increase the generalizability of the theory or the hypothesis. Quantitative researchers may also use inductive reasoning as they look for similar experiences and results and form new ideas, concepts, or theories.

Basic Versus Applied Research Approaches

Research often strikes many students as abstract and distant from real life. This is especially true if research aims primarily at knowledge creation through theory building. The goal of **basic research** is to design studies that can test, refine, modify, or develop theories. As an example of basic research, Marcia's (1966)

TABLE 1.2 Qualitative and Quantitative Approaches: The Scientific Process

The Scientific Process	Qualitative Research	Quantitative Research
Step 1: Ask a general question	Observation, reflection, and a review of research leads to a question	Review of research and theory leads to a question
Step 2: Generate more specific questions or a research hypothesis	Inductive reasoning leads to more specific yet flexible research questions	Deductive reasoning leads to a research hypothesis
Step 3: Collect data to answer question or hypothesis	Data are in narrative or image form, collected through methods such as interviews or observations	Numerical data collected, such as tests, checklists, surveys
Step 4: Data analysis	Identify patterns or themes	Conduct statistical analysis
Step 5: Interpret finding	Make conclusions based upon themes and patterns	Hypothesis rejected or accepted based upon statistical results

research on adolescent identity led to a refinement of one stage of Erik Erikson's psychosocial theory of development. Marcia's goal was not to create a program to address practical ways to help adolescents but rather to extend and support the theory. In contrast, applied research does try to apply knowledge to actual practice. **Applied research** studies examine the effectiveness and usefulness of particular educational practices. Here the goal is to determine the applicability of educational theory and principles by testing hypotheses within specific settings. For example, Schmitt-Rodermund and Vondracek (1998) examined whether parenting behaviors predicted the amount of adolescent identity exploration as described by Marcia. The results of their study have implications for how parents and adolescents interact.

Both basic and applied methods of research have their places in the educational research field. To some degree, the approach selected depends on whether the findings are utilized and result in a change in practice. In basic research, the overarching goal is to develop and modify theory. These theory-based studies, while critical to the formulation of applied research, often have low utilization and do not result in systemwide change. Although the goal of applied research is to demonstrate the usefulness of theories in practice, the reality is that applied research studies often take many years to stimulate change, even when the findings are disseminated to large groups of individuals through applied research journals. Two approaches that do result in more immediate change are program evaluation and action research.

PHILOSOPHICAL FRAMEWORKS FOR EDUCATIONAL RESEARCH

Educational research today is beginning to move away from a hard and fast distinction between qualitative and quantitative research methods. In fact, many researchers combine both approaches in order to gather a breadth of data and to validate their results. Researchers can, however, be separated into groups based on their philosophical frameworks, identified by the assumptions they make about the nature of the reality being studied, claims about what we can and cannot know, and the ways in which they utilize theories and findings. Each framework makes assumptions about whether qualitative or quantitative methods are most appropriate for extending our knowledge about education. As a beginning researcher, it is important that you consider which approach best captures your own assumptions about how the world works.

Scientific Realism

Scientific realism is a term applied to the framework used by most researchers who take a purely quantitative approach to research. Quantitative research is characterized by a desire to answer research questions by producing numerical data that represent various constructs and variables. A **construct** is a hypothetical concept that is typically developed from a theoretical framework. Although constructs are names for things that cannot be seen (for example, intelligence, motivation, self-esteem), they are assumed to be real characteristics that influence educational outcomes. When constructs are measured in educational research, they are known as **variables**. Like the constructs they represent, variables are defined as attributes, qualities, and characteristics of persons, groups, settings, or institutions, such as gender, social skills, socioeconomic status, exclusiveness, or achievement. Scientific realists strive to establish cause-and-effect relationships when possible, using data collection methods such as questionnaires, tests, and observational checklists to produce quantitative data.

The philosophical underpinnings of the scientific realism approach can be found in the arguments of philosophers known as positivists who have primarily tried to describe knowledge generation in the physical sciences. The first assumption made by scientific realists is that there is a real social and psychological world that can be accurately captured through research. In other words, there is an objective reality that research aims to describe. Scientific realists further assume that the social and psychological world can be studied in much the same way as the natural world, by breaking complex phenomena and problems into smaller parts. The major job for the researcher is to identify the most important parts or variables and accurately describe how these are related to each other in the real world. However, because humans are fallible and social scientists study human characteristics, reporting that reality must be done with a certain degree of probability. Scientific realists see knowledge as conjectural (Phillips & Burbules, 2000) and therefore subject to possible revision. All hypotheses are tested using statistical tests that establish the level of confidence that one can have in the results obtained. Scientific realists do recognize that because educators study human behaviors and characteristics, research may be influenced by the investigator. For an investigator to maintain clear objectivity, he or she must play a detached role through which there is little opportunity for interaction with the participants under study. Scientific realists believe that inquiry can be value-free and that a researcher who strives to eliminate any personal bias can reliably determine findings. Although they borrow rigorous scientific techniques from the natural sciences, they recognize that, in education and psychology, true scientific experiments are not always possible. Scientific realists concede that different persons might have different perceptions of reality; however, they assume that experiences overlap to a large degree and that a good researcher can take these different perceptions into account in providing the best possible explanation of reality.

Social Constructivism

Traditionally, purely qualitative research is often done by persons who hold a framework referred to as *interpretive*, *constructivist*, or *naturalistic*. (We use the term **social constructivism** to refer to this approach.) Social constructivists challenge the scientific realist assumption that reality can be reduced to its component parts. Instead, they argue that phenomena must be understood as complex "wholes" that are inextricably bound up with the historical, socioeconomic, and cultural contexts in which they are embedded. Therefore, social constructivists attempt to understand social phenomena from a context-specific perspective.

Social constructivists view scientific inquiry as value-bound and not value-free. According to Lincoln and Guba (1985), this means that the process of inquiry is influenced by the researcher and by the context under study. This philosophical perspective argues that reality is socially constructed by individuals and this social construction leads to multiple meanings. Different persons may bring different conceptual frameworks to a situation based on their experiences, and this influences what they perceive in a particular situation. In other words, there is no one true reality, nor can one assume that the experiences that people have had will overlap to a large degree. Rather, we construct reality in accord with the concepts most appropriate to our personal experiences. Therefore, the researcher must attempt to understand the complex and often multiple realities from the perspectives of the participants. The acceptance of the existence of multiple realities leads social constructivists to insist that a set of initial questions asked in a study will likely change or be modified as these multiple realities are uncovered or reconstructed during the process of conducting research. The only true way to accomplish this understanding is for the researcher to become involved in the reality of the participants and interact with them in deeply meaningful ways. This provides an opportunity for mutual influence and allows the researcher to see the world through the eyes of the participants. "The inquirer and the object of inquiry interact to influence one another; knower and known are inseparable" (Lincoln & Guba, p. 37). This approach, then, requires that researchers use data collection methods that bring them closer to the participants using techniques such as in-depth observations, life histories, interviews, videos, and pictures.

Advocacy-Liberatory Framework

Researchers taking an **advocacy-liberatory** framework for research also assume that there are multiple possible realities that are dependent on social, political, and economic contexts. However, these researchers go beyond the social constructivist claim that researchers' values can influence research by insisting that moral values should form the impetus for research and that research should seek to improve the lives of persons who have little social power and have been marginalized by more powerful groups in their societies. In essence, the goal of advocacy or liberatory researchers is liberation through knowledge gathering. Paulo Freire (1921–1997), a literacy worker from South America and author of *Pedagogy of the Oppressed* (1970), based his philosophy of research on these principles and argued that research should provide freedom from oppression and debilitating living environments. Working on literacy skills with poor and oppressed Chilean workers in the 1960s and 1970s, Freire asserted that research should be conducted in a collaborative manner, with community members participating in the selection and analysis of themes during data analysis. This collaboration requires that the researcher engage in respectful dialogue with the study participants and understand reality from the perspectives of the community. According to Freire and other advocacy-liberatory investigators, research should not only use inductive processes to gather information but engage in research as a form of social advocacy in which participants identify the types of changes sought. Whereas this type of research usually uses qualitative methods of data collection, it might use quantitative methods constructed in collaboration with participants if these data will help the people achieve social changes in their society. The type of data collected is less dependent on philosophical assumptions than by its potential to illuminate experiences and facilitate action to achieve a better life. In other words, research should be used not only to educate and produce knowledge but also to empower people to take political action and use their political voice to change and improve their place in society.

Pragmatism

Pragmatism is the framework that has been most developed by American philosophers. Unlike the other frameworks, pragmatism is not concerned with whether research is describing either a real or socially constructed world. Instead, for pragmatists, research simply helps us to identify what works. Of course, we might ask our pragmatists what they mean by what works. They are likely to reply that knowledge arises from examining problems and determining what works in a particular situation. It does not matter if there is a single reality or multiple realities, as long as we discover answers that help us do things that we want to do. A pragmatist might insist that a good theory is one that helps us accomplish a specific goal (or set of goals) or one that reduces our doubt about the outcome of a given action. Most pragmatic researchers use a mixed-methods approach to research; for example, they use both qualitative and quantitative methods to answer their research questions. Pragmatic researchers propose that even within the same study, quantitative and qualitative methods can be combined in creative ways to more fully answer research questions. Campbell and Fiske (1959) are often thought to be among the first researchers to introduce the notion of using both qualitative and quantitative techniques to study the same phenomena. In current research, pragmatic frameworks are used by both professional researchers and researchers who are primarily practitioners (for example, teachers, counselors, administrators, school psychologists).

The assumptions underlying the philosophical frameworks described previously are summarized in Table 1.3.

RESEARCH ETHICS

Regardless of the type of research conducted, research ethics is an important consideration. Most professional organizations have their own codes of ethics (see the American Psychological Association and the American Sociological Association for examples). In addition, colleges, universities, and other institutions that conduct research have **institutional review boards** (IRBs) whose members review proposals for research to determine if ethical issues have been considered. If you are conducting research in a noncollege setting, in an elementary or secondary school or a community organization, there may not be a committee called an "IRB." In this case, you will need to find out who will review your proposal and the procedures you will need to follow to obtain approval.

TABLE 1.3 Frameworks and Assumptions Underlying Educational Research

	Scientific Realism	Social Constructivism
Knowledge-oriented approaches	• Research aims to describe an objective reality that most or all people would agree is real • Educational settings and problems can be studied by empirical analysis of component parts • Research should be value-free • Researchers should be detached from participants and strive to be objective • Theories and hypotheses are formed and then confirmed or disconfirmed through collection of data	• Reality is historically and culturally constructed so there are multiple possible realities • Educational settings and problems must be understood as complex wholes • Researchers must continually strive to be aware of and control their values • Researchers should become actively involved with participants in order to understand their perspectives • Theories and hypotheses are generated during data collection and achieve meaning through human interactions
	Advocacy-Liberatory	**Pragmatism**
Action-oriented approaches	• Reality is socially constructed and influenced by social, political, and cultural inequalities • Although qualitative methods are preferred, educational settings and problems can be studied using any methods that truly represent the experiences of the participants • Research must be based in values and should empower marginalized groups to improve their lives • Researchers should collaborate with participants as equal partners • Theories and hypotheses should provide action plans to achieve a better life	• The immediate reality of solving educational problems should be the focus of educational research • Educational settings and problems can be studied using any method that accurately describes or solves a problem • Research should strive to find ways to make education better • Researchers should collaborate with participants to fully understand what works • Theories and hypotheses are useful tools in helping to improve education

For the most part, issues of ethics focus on establishing safeguards that will protect the rights of the participants. The traditional and often dominant issues that emerge when considering research ethics involve obtaining **informed consent** from participants, protecting them from harm, and ensuring **confidentiality**. Informed consent means that participants have been given information about procedures and risks involved in the study and have been informed that their participation is voluntary and they have the right to withdraw from the study without repercussions. IRB committees typically scrutinize research proposals for these issues and will weigh any potential risk to the participants against any possible gains for science. Keep in mind that the process for addressing ethical issues might change the participants in your study. This may happen because some of the people that you have selected will not agree to be in the study, or the IRB may not give you permission to use those participants. Even if you use a random sampling procedure, the final participants in your study are volunteers. All quantitative researchers should consider how this might change the representativeness of the sample or exclude key informants.

The members of the IRB review proposals and examine the methods described in the proposal to ensure that all ethical considerations have been addressed and that sufficient detail of the actions to be taken by the researcher are provided. As already mentioned, most institutions whose students, professors, and staff conduct research have their own IRB committees that provide specific guidelines for study approval. IRB committees are mandated by national legislation (National Research Act, Public Law 93–438). To be well versed in specific requirements and procedures, you should contact your own university or college's IRB committee. IRB committees typically require that the researcher prepare a document that includes the following:

- *A cover page.* In this the researcher introduces the principal investigator and his or her qualifications and contact information, the project title, and the type of research that is being proposed.
- *A detailed description of the study*. This includes a summary of the literature, the research method, the significance of the research, and particulars about the location and duration of the study. The committee will want the specifics on any treatment (for experimental research) and any instruments or protocols that might be used.
- *A description of the participants*. The researcher needs to include background information on the individuals in the sample and the sampling procedures to be used. If the participants are to be selected from a specific institution (school, hospital, club, and so on), then written permission is needed from the director or principal.

- *Discussion of inducements, benefits, or compensation to be offered to the participants*. If the participants are provided any incentives to encourage their participation in the study, the specifics should be explained.
- *Analysis of the risks and benefits*. If there are any potential risks to the participants, the researcher needs to clearly indicate how the benefits outweigh any risks.
- *Informed consent, confidentiality, voluntary nature of the project, and debriefing activities.* A description of the procedures for obtaining informed consent must be described in order to obtain IRB approval. This is particularly important when children younger than eighteen years are participants. In this case, parental consent must be obtained. Most IRB committees require that the researcher submit copies of the informed consent forms. This consent form must include at least the following information:
    - Detailed description of the project
    - Description of any potential risks involved
    - The voluntary nature of the study
    - A confidentiality statement

Some educational studies might be considered IRB exempt. These are studies that involve commonly accepted educational practices and those that report the anonymous results of educational tests (such as those published on Web sites). For example, new instructional practices or program evaluations commissioned by the school to refine ongoing practice would be exempt. You must consult your IRB committee if you believe that your study falls into this category. Some IRB committees also have expedited forms that can be used under specific circumstances. Once again, to determine the type of IRB form that is necessary for your study, you should contact your IRB committee. Most important, the IRB documents must be submitted and approved before any data are collected. So be sure to start this process early.

Keep in mind that the procedures outlined in one's IRB proposal are not cast in stone, meaning that they can be modified to meet the ever-changing needs of the researcher and the overall study and its purpose. This, in fact, is common. In many situations when researchers enter the field and start conducting their study, they find that the real-world situations vary considerably from what they had envisioned when creating their proposal on paper. Minor changes to a proposal method typically do not require that one go back and be reapproved by the committee or board that initially reviewed and approved the study. Some examples of these modifications might include increasing the study's sample size or modifying items on the survey following its pilot. Because of its exploratory nature, modifications are expected in qualitative research. When conducting a case study or ethnographic study, it may be necessary to submit a series of small research proposals to the committee instead of one large proposal. This will allow the qualitative researcher to start out with broad procedures when entering the research field and to maintain the ability to refine the process as the study progresses.

Ethical issues pertinent to each type of research will be discussed in chapters 6, 8, 9, 10, and 12 on the specific types of research.

SUMMARY

While educational accountability is not new to American education, NCLB has led to many new changes in schools and school districts. This legislation has increased the visibility of test scores and has imposed regulations forcing school districts to demonstrate their AYP. Recent research on the effectiveness of NCLB to improve test scores has increased the debate regarding the future of the legislation. Some districts have moved beyond APY and use value-added assessment to demonstrate academic progress.

In the light of reform efforts and increased responsibility for policy- and decision making, educational professionals are turning to research to solve educational problems. There are many different approaches and techniques that might be used to improve the quality of teaching and learning. Despite the narrow definition of "reliable" research embodied in the NCLB legislation, you can now see that multiple methods are available for investigating issues in education. These methods utilize the scientific method as an underlying framework.

Whereas NCLB focuses on quantitative research both quantitative and qualitative approaches utilize the scientific method but in different ways. These approaches have grown out of differing philosophical views, but all focus on systematically endeavoring to answer questions about what works in education, and therefore, we argue, all have value in the ongoing debate about improving education. The research approaches we have described have evolved out of philosophical frameworks that demonstrate widely divergent views on the nature of reality, how we come to know that reality, and whether we need to concern ourselves with such ponderings as long as we can figure out what works and what does not! The philosophical viewpoint to which you subscribe will, in many ways, determine what research you are willing to undertake or accept as meaningful.

Practitioners who undertake research often turn to action research as a research design. Action research is focused on practical problems and has as its primary goal to improve educational practice. This method often gathers both qualitative and quantitative data.

Any research, regardless of the type, must first be ethical. The ethics of a research study are typically determined by an institutional review board (IRB). The IRB will evaluate if the study meets the basic requirements. The participants must:

- Understand the voluntary nature of the study
- Sign a consent form
- Be guaranteed confidentiality, and
- Be debriefed of informed consent

As you can see, there are many issues that must be considered prior to the implementation of an educational research study. As a professional in the field of education, you will be challenged on a daily basis to create and sustain an effective learning environment. An understanding of educational research and its philosophical underpinnings is vital to making informed decisions about what research you will use to support your everyday practice.

KEY CONCEPTS

action research
adequate yearly progress (AYP)
advocacy-liberatory research
applied research
basic research
confidentiality
construct
deductive reasoning
hypothesis
hypothetic-deductive method
inductive reasoning
informed consent
institutional review board (IRB)
pragmatism
qualitative research
quantitative research
scientific realism
scientific research
schools in need of improvement (SINI)
social constructivism
theory
value-added assessment
variable

DISCUSSION QUESTIONS OR ACTIVITIES

1. NCLB and the accountability movement emphasize testing students and setting required benchmarks for student progress. Explore the tests and benchmarks used in your state or school district and discuss the factors that might influence whether schools meet their mandated goals.
2. Pick a philosophical framework that is closest to your personal belief about how knowledge is generated, and find a student in your class whose preferred framework is different from yours. Debate the pros and cons of each framework as a guide to research.
3. Pick an educational problem or topic, and discuss how it could be explored using one quantitative and one qualitative approach.

4. Discuss the essential differences between the types of assessment used under NCLB and value-added assessment. Engage in a debate on the pros and cons of each approach.
5. Discuss the value of action research. Generate an example of an action research study that might be conducted in your field.

SUGGESTED READINGS

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher, 27*(2), 65–75.

Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math trends.* Cambridge, MA: Civil Rights Project at Harvard University (The). (ERIC Document Reproduction Service No. ED491807)

Onwuegbuzie, A. J. (2000, November). *Positivists, post-positivists, post-structuralists, and post-modernists: Why can't we all get along! Towards a framework for unifying research paradigms.* Paper presented at the Annual Meeting of the Association for the Advancement of Educational Research, Ponte Vedra, FL.

Peterson, P. E., & West, M. R. (Eds.). (2003). *No child left behind? The politics and practice of school accountability.* Washington, DC: Brookings Institute.

Popham, J. W. (2004). *America's "failing" schools: How parents and teachers can cope with No Child Left Behind.* New York: RoutledgeFalmer.

CHAPTER TWO

TYPES OF EDUCATIONAL RESEARCH DESIGNS AND RELATED MAJOR CONCEPTS

CHAPTER OBJECTIVES

- Explain the differences between and provide an example of qualitative and quantitative research
- For each research method discussed describe the key elements of that approach and provide an example of a research study that uses each approach
- Be able to explain all of the major concepts associated with each approach

TYPES OF APPROACHES USED IN EDUCATIONAL RESEARCH

The philosophical frameworks described in chapter 1 provide the underlying assumptions associated with the different research designs discussed in this chapter. To some extent, your philosophical framework will guide your selection of the type of research approach you will use. As already discussed, the types of research in education are often described along the basic-to-applied continuum. Research designs are often further classified by (1) the methods used to design the study and to collect data (for example, qualitative versus quantitative approaches), and (2) how the information is shared (for example, the dissemination of the findings).

These approaches are discussed in detail in later chapters. Here we provide you with "executive summaries" and a review of major concepts associated with each approach. For many of you, a research proposal (see Appendix A for details) that necessitates that you read research conducted in your field of study is required in this course. It is our hope that these summaries, along with the major concepts associated with each research design, will help you to begin to read and critique research for your proposal. Furthermore, a basic familiarity with these designs may assist you in determining the type of research used in the study you are reading.

To facilitate your understanding of the different research designs used in education, we divide the approaches into three categories: quantitative approaches, qualitative approaches, and approaches that incorporate both.

Quantitative Research Approaches

Simply put, all quantitative research approaches summarize results numerically. This research draws on a principle of scientific realism: there is a single reality that can be described by numbers. While all quantitative approaches collect and analyze numbers, quantitative research is often classified as either **nonexperimental** or **experimental**. Experimental research is designed to determine cause-effect relationships. Nonexperimental research uses numbers to describe preexisting groups or to determine whether a relationship exists between variables.

For beginning researchers, there are many important concepts to understand related to quantitative research. Since we mention variables in the previous paragraph, let's start with a definition of variables. A **variable** is a characteristic or attribute that varies! This definition is the ultimate circular definition. As is, it does not tell us much. Let us back up and consider a few important elements that will help us to understand the term *variable*.

In education and psychology, we are often confronted with **constructs**. A construct is a hypothetical concept that cannot be directly observed. For example, intelligence is a construct. We hypothesize that intelligence exists, but we cannot see it. We cannot surgically open up someone's brain and locate intelligence. However, based on theory, *intelligence* is a word used to explain why people differ in their abilities to reason and solve problems. Reasoning and problem solving are behaviors that can be measured, and so are referred to as variables. So even though we cannot directly measure a construct such as intelligence, we can measure the variables that are theoretically related to it. Therefore the researcher creates an **operational definition** that specifies how the variables will be measured. So if we operationally define intelligence as a score on the Stanford-Binet Intelligence Test, other researchers understand that in our study, intelligence is conceptualized as the ability to successfully complete the items on the test. Of course, operational definitions are limited, and the construct of intelligence is much broader and more complex. Therefore, some researchers might also include a more conceptual definition of intelligence, such as: intelligence is the ability to solve complex problems. Operational definitions may seem like a strange way to define a variable. However, they help to clarify exactly how the concept is being measured or used in a study. If two researchers use different tests to measure intelligence, we can look at their tests to see how they might define intelligence in similar or in different ways.

While quantitative researchers study variables with a particular group of participants, their hope is to use their findings to draw conclusions beyond those participants. To accomplish this, quantitative researchers must select their participants in a systematic way. A sample is selected from a larger population through **random selection**. A **sample** is a smaller version of the **population,** the group to which the researcher would ultimately like to generalize or apply the results of the study. In order to make these generalizations, the researcher should select the sample using a random selection technique. Random selection is a procedure in which each and every member of a population has an equal and independent chance of being selected for a sample. Random selection is used so that the sample can be representative of the population from which the sample is taken. Failure to use randomization can result in **sampling bias**, a situation in which the sample does not accurately reflect the population. For example, a researcher wants to know the extent of cyberbullying among middle school girls. She recruits her sample by posting an invitation to take part in an online survey on several social networking Web sites. One hundred girls complete the survey; the researcher finds that 75% report that they have been cyberbullied at least once. Can she conclude that this is an accurate estimate of cyberbullying in the overall population of middle school girls? Possibly not, because the sample may not include girls who are bullied through cell phone contact but do not participate in social networks. So the figure of 75% might be too high or too low. Perhaps girls who volunteer to take the survey are more likely to have been cyberbullied than girls in the overall population. To obtain a more representative sample, the researcher should consider randomly selecting middle schools from a defined geographic area and then surveying all girls in those schools.

With an understanding of some of the important concepts related to quantitative research, let's explore the unique features of different quantitative approaches.

Descriptive Survey Research Descriptive survey research aims to describe behaviors and to gather people's perceptions, opinions, attitudes, and beliefs about a current issue in education. These descriptions are then summarized by reporting the number or percentage of persons reporting each response. The survey is the primary method used to gather such data or information from people. Although more and more technology and Web-based surveys are being used in research, the long-standing paper-and-pencil survey continues to be the main mode of data collection. A commonly held misconception is that descriptive survey research is an easy method, requiring simple questions and answers. This just is not so. Good descriptive survey research requires thoughtful and careful planning.

Planning to conduct survey research begins with a clear research question. For example, let's say that a researcher is interested in gathering educational administration students' opinions regarding their graduate school training. The research question "What are educational administration students' opinions regarding their graduate school training?" would form the framework of the survey. However, you will note, this is a rather broad question. In order to gather breadth and depth of data, the researcher would want to further analyze this broad question and develop a set of **subquestions** in order to develop the actual survey. Subquestions, in survey research, are more specific, more targeted questions and assist the researcher in the development of the actual survey questions. Researchers then use the subquestions as well as the literature they have read on the topic to develop their survey items. In our example, a researcher might consider the following subquestions:

1. What are students' opinions regarding their relationship with their professors?
2. What do students think about their level of preparation regarding assessment?
3. How do students view their ability to do research?
4. How do students view their ability to negotiate the job market?

The survey items would be developed from these four subquestions.

Once the survey is developed, the researcher carefully analyzes the items and conducts a **pilot study**. A pilot study involves the administration of the survey to a small group of individuals who help to work the "kinks" out of the survey. This process helps the researcher determine the survey's **validity** and its **reliability.** Briefly, the validity of a survey is the degree to which the survey measures what it was intended to measure. If a researcher carefully constructs the survey items to match the subquestions, the survey should be valid (process is described in more detail in chapter 8). For example, using subquestion number 1, regarding students' relationships with their professors, survey items might include the following:

- To what extent were your professors responsive to your concerns regarding the amount of work assigned?
- To what extent were your professors willing to discuss any issues that developed during the course?
- To what extent did you feel comfortable discussing your opinions with your professors?

The survey items in this example are closely related to the subquestion regarding relationships with professors. The researcher is well on her way to establishing validity. Reliability of a survey refers to the consistency of responses. That is, if the survey is administered to the same group of individuals more than once, would the responses be the same? Survey researchers use many techniques to enhance the reliability of their surveys. These techniques are discussed in detail in chapter 8, on descriptive survey research.

Survey researchers may also include **demographic items** on their surveys. Demographic items are designed to obtain background information on participants. This type of information allows the researcher to report the findings in a context such that the results can be better interpreted.

Once reliability and validity of the survey have been established, the researcher is ready to send the survey to the target group. In many cases, the researcher will use a random selection procedure in order to identify possible participants for the survey. The survey will be given to a sample and, if the sample is large and representative enough, the results generalized to a population.

While an estimated 70% of research in education is survey research, it is not without its problems. One concern has to do with **response rates**. Response rates are the number of individuals who respond to the survey. Typically, in survey research the response rates are low, jeopardizing the researchers' ability to draw conclusions and generalize the results back to the population.

Experimental Research For decades, experimental research has been considered the strongest quantitative approach for drawing conclusions about cause and effect. Indeed, the manner in which NCLB defines *scientific research* has resulted in an increased use of this method. Often when people think about research and what research is, they commonly associate it with characteristics typical of experimental research. The most common purpose of experimental research is to determine whether a particular approach or way of doing something different is "better" than a more traditional approach that has served as the standard practice. (Keep in mind that sometimes experimental research is conducted with hopes that no difference will be found between the two methods or approaches under investigation.) Experimental research studies the effect or the impact of an approach under stringent and controlled conditions to make statements of causality. Sometimes, these conditions involve random selection of a sample from the larger population. The randomly selected participants constitute the sample. Participants in the sample are then randomly assigned to one of two or more groups that are treated (manipulated) with regard to a specific educational approach or practice or are exposed to different treatments or no treatment (the **control group**). In some studies, specifically repeated measures studies, the same group is given the treatment at different points in time. In both situations, the manipulated or experimental treatment conditions are called the **independent variable**, which precedes and is assumed to cause a change in the measure referred to as the **dependent variable**. For example, a researcher might ask, "Does instructional strategy in reading (a skills-based approach or a literature approach, or an approach that combines the two) affect reading achievement of fourth graders?" In this study, reading achievement would be the dependent variable, and instructional strategy would be the independent variable. The sample would be a group of fourth graders randomly selected from the population to which the researcher wants to generalize the study results (for example, fourth graders in an entire school district). Students would be randomly assigned to receive either a skills-based or literature approach, and reading achievement would be measured.

The final component of an experimental study is to control **extraneous variables**. An extraneous variable is any variable other than the independent variable that might influence the dependent variable. In any experimental study, there are many possible extraneous variables. In the study just described, one would need to consider whether the teachers for each class were equally good teachers. The amount of time spent on reading instruction might also be an extraneous variable. Differences in student abilities before the instruction begins are an extraneous variable that is controlled through random assignment.

Independent variables are typically **nominal** or **categorical** in nature. For example, type of intervention (ABC intervention versus no ABC intervention) is a categorical variable because an individual is assigned to one group or the other and may be a member of only one group. Dependent variables are often **continuous variables.** These variables can take on several different levels. Consider the variable reading achievement which is a continuous variable because scores on a reading achievement test can assume many different values.

Once the variables are clearly identified and defined, the researcher conducts a review of related literature and develops a research hypothesis. A **hypothesis** is a conjectural statement regarding the expected outcome of a research study. Simply put, it is an educated guess based upon a review of the research conducted on the research problem or the variables that are being investigated. Hypotheses are tested through the use of statistical techniques that are discussed in greater detail in chapter 3.

When designing their studies, experimental researchers are concerned with experimental validity because of their desire to draw cause-and-effect conclusions. In order to strengthen their legitimacy of doing so, an experimental researcher designs a study that controls for **threats to internal and external validity**. A study is considered internally valid when the results or outcome of the study can be attributed to the experimental treatment and not an extraneous variable. Let's say in the study on reading described earlier, the group that received the combined instructional approach, for example, both the skills-based and literature approaches, performs better than the groups who received a single instructional approach. In this case, the researcher wants to attribute the outcome to the use of the combined approach or the treatment effect. Researchers can accomplish this by setting up a well-controlled study. **Random assignment** to treatment and control conditions is one way to control for threats to internal validity. Other mechanisms for control are discussed in chapter 9, on experimental research.

In addition to internal validity, there are threats to external validity. External validity is the extent to which the results of the study are generalizable. Remember, most experimental researchers want to generalize the results of the study back to the population from which the sample was selected. There are many ways to ensure the generalizability of a study, but once again random selection plays a key role.

It should be noted here that random selection and assignment of participants to treatment and control groups is something not easily achieved when one is doing research in a school. What more typically happens in a school is that the researcher is given permission to use groups (for example, classrooms) that have already been established. Under these conditions, the researcher cannot randomly assign the participants to different groups or treatment conditions. By using already established groups, the researcher is conducting a **quasi-experimental research**. Quasi-experimental research is a form of experimental research in which the researcher does not have control over assignment of individuals to conditions but can randomly assign whole groups to different treatments. Typically groups are pretested prior to the treatment to be sure they do not differ on the dependent variable. Let's say a researcher studying mathematics achievement is given permission to study the classrooms of two teachers from the same school. The researcher is interested in determining if the use of graphing calculators affects student understanding of slope of lines. Neither class has previously used graphing calculators. Both teachers agree to assess students' initial understanding of the geometric concepts underlying slope. The results of this assessment indicate that the average scores of the two classes on this pretest are similar. The researcher then flips a coin to see which group will get the graphing calculators. Both teachers spend the same amount of time teaching the concept of slope. At the end of one week, students are tested again to determine if the use of the graphing calculators produced better understanding than learning with no calculators.

Both experimental and quasi-experimental designs are accepted under the conditions for scientifically based research set forth under NCLB.

Single-Subject Research Single-subject research is a special type of experimental research. Unlike the experimental designs explained earlier that make group comparisons, single-subject research examines treatment effects on individuals or a small group of individuals. In this type of research, which is often used by teachers or professionals working with special needs students, a treatment or an intervention is manipulated. The famous behavioral psychologist B. F. Skinner used single-subject research that examined individual change over time in response to a particular reinforcement condition.

The critical part of single-subject research involves the researcher collecting several sets of data prior to the implementation of the treatment. This is called a **baseline phase** because data are collected at multiple points throughout a period of time. It establishes the person's performance level before a treatment is given. Then a treatment is introduced during the **treatment phase**. Again multiple measures are taken and compared to baseline data to see if the treatment changes the behavior. For example, a school psychologist is working with a child who has been repeatedly involved in self-stimulatory behaviors. The child has a ball of string that he continues to wind and unwind. The school psychologist observes the student in class and counts the number of times the child engages in this behavior over a two-week period (the baseline). After consultation with the teacher, the literature on best practice, and the child's parents, the school psychologist decides that physical activity would be an appropriate intervention and would be incompatible with the self-stimulatory behavior. During the treatment phase—the physical activity—the school psychologist continues to monitor the number of times the child uses the ball of string. In order to determine whether the treatment is effective, the school psychologist compares the levels of self-stimulatory behavior during the baseline and treatment phases.

In some single-subject research studies, the treatment is withdrawn, and in others, the treatment continues. Many factors influence the type of single subject design a researcher chooses. The choice whether to withdraw the treatment depends upon the nature of the behavior, the child, and the nature of the treatment.

Causal-Comparative Research Causal-comparative research, or **ex post facto** (after-the-fact) **research,** is a research approach that seeks to explain differences between groups by examining differences in the experiences of group members. This design is nonexperimental. However, like experimental research, it examines the effect of an independent variable (the past experience) on a dependent variable while also trying to control extraneous variables. However, unlike experimental research, the independent variable (the past experience) has either already occurred, or it would be unethical to manipulate. Unlike experimental research, the researcher does not have control over the independent variables, making statements of causality more difficult. When examining causal-comparative research, the results are suggestive of possible causal relationships but clear cause-and-effect statements should be avoided.

For example, let us say that you are interested in what causes the differences in the readiness skills of kindergarten students. After reading past research studies, you decide to examine preschool attendance as an independent variable that might have "caused" a difference in kindergarten readiness (the dependent variable). Preschool attendance has already occurred or happened; as a researcher, you cannot control or manipulate it. If you were to conduct such a study, you would simply identify two groups, one group that attended preschool and one group that did not, and then measure and compare school readiness scores. If the groups differ on their readiness scores, you infer that preschool attendance may have caused the readiness scores to differ. However, caution is warranted. Because no random assignment occurred, the two groups being studied could be very different to begin with, which might mean that other factors (extraneous variables) caused the difference in readiness scores. For example, there may be differences in family income or parental levels of education (or both). Therefore, making sure that the two comparison groups are as similar as possible on all other extraneous variables (other than the independent variable) is a critical part of designing a causal-comparative study. There are many ways to control for extraneous variables.

(These techniques are discussed in greater detail in chapter 9.) One way you could control for the extraneous variable *parental level of education* would be to include only those students whose parents have a college education or higher in the research study.

Correlational Research Another quantitative method that is considered nonexperimental is **correlational research**. This type of research seeks to determine the degree of relationship between two or more variables. Correlational researchers are interested in studying only one group of individuals (for example, fifth grade students) on two or more variables that are not manipulated or controlled by the researcher (for example, reading scores and IQ). The variables are examined to determine whether they are related and, if so, the direction and magnitude of that relationship. Traditional educational research textbooks argue that simple correlational research cannot demonstrate that one variable is causing a change to occur in another variable. Rather, the main purpose of correlational research is to determine, through application of a quantitative statistical analysis, whether a relationship exists between the variables under investigation. One might make predictions based on these relationships, but not statements of causality. For example, if such a relationship does exist, the strength and the direction of the relationship are reported numerically in what is referred to as a **correlation coefficient**. Scores from this analysis produce a correlation coefficient that ranges from negative 1.00 to positive 1.00. Note that negative and positive do not have any "moral value" attached to them in this context. A highly **negative correlation** is not a relationship that is bad but one that results from scores on two variables moving in opposite directions: an increase in one variable is accompanied by a decrease in the other variable being studied. For example, as absentee rates increase, student achievement decreases. A **positive correlation** is a relationship where both variables move in the same direction. That is if one variable increases, the other variable increases. If one variable decreases, the other variable decreases. If a researcher finds that as creativity increases, intelligence also increases a positive relationship would be reported.

Many beginning research students, as well as seasoned researchers, often confuse causal comparative and correlational research. There are in fact similarities between the two types of research. Both are nonexperimental and do not involve the manipulation of an independent variable. Furthermore, statements about causality cannot be firmly verified using such studies. So how are they different? The traditional answer is that correlational research examines relationships in only one group, while causal comparative research examines differences among two or more groups. However, consider the following studies. Researcher A is interested in conducting a study that examines whether number of hours participating in after-school mentoring programs influences student achievement. She considers the number of hours spent in mentoring a continuous variable (for example, 10–60 hours). In this case, the study is correlational in that the study is investigating the relationships between the variable mentoring and the variable achievement. Researcher B is also interested in the same study. However, this researcher decides to categorize number of hours into three different groups (10–20 hours, 21–40 hours, 41–60 hours). In this case, the study becomes causal comparative. Let's say both studies demonstrate that the longer a child spends in a mentoring relationship, the higher the achievement level. Traditional researchers might argue that only in the case of the causal comparative study can we even begin to discuss a possible causal connection between mentoring and achievement. In reality, with no controls discussed in either case, neither can make any statements about causality! So how different are these two designs? Some researchers argue that there is little difference between correlational and causal comparative research. We continue this debate in later chapters.

Meta-Analysis As you begin to conduct research for your research proposals, you may find studies that fall under the category of meta-analysis. Some of the unique features of meta-analysis tend to pose a dilemma for students new to the area of research. Meta-analysis studies ask a research question and then conduct a review of literature to find other studies that have researched the same question. The meta-analysis researchers do not collect any new data of their own to answer the research questions. They statistically summarize the results of other studies and use their analysis to make conclusions about the research question. Now perhaps you see why so many consider it to be confusing. The purpose of a meta-analysis is to ask a research question and use past quantitative studies as data to answer the question. The data from these studies are reanalyzed using an appropriate statistical analysis, and a typical result, usually referred to as an **effect size**, across all studies is reported. Effect size usually provides an estimate of how powerful a given treatment was across the studies examined.

Qualitative Research Approaches

According to Denzin and Lincoln, as cited in Denzin and Lincoln (2005):

> Qualitative research is a situated activity that locates the observer in the world. It consists of a set of interpretive, material practices that make the world visible. These practices transform the world. They turn the world into a series of

> representations, including field notes, interviews, conversations, photographs, recordings, and memos to the self. At this level, qualitative research involves an interpretive, naturalistic approach to the world. This means that qualitative researchers study things in their natural settings, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them. (p. 3)

There are many important elements in Denzin and Lincoln's definition. First, the definition places the role of the researcher in the position of an observer. That is, in order to uncover reality, the researcher must use the power of observation to clarify and give meaning to the phenomena being studied. Second, Denzin and Lincoln argue that qualitative researchers have the opportunity to stimulate change in the world (or the group being studied). Becoming agents for change is similar to the set of beliefs put forth by those who embrace the advocacy-liberatory and social constructivist philosophical framework discussed in chapter 1.

Qualitative researchers develop a broad research question or area of focus based upon their observations, readings, or experiences in the research setting. Additionally, the qualitative researcher develops specific questions to help guide his or her observations. These questions are also referred to as qualitative subquestions or in some qualitative literature as foreshadowed questions. It is important to note that these questions may change as the study progresses. Unlike their quantitative counterparts whose research question remains constant throughout the research, qualitative research questions are more flexible. They may use their early experiences in the research setting to modify or expand their research questions.

Qualitative research techniques, beyond observation, bring the researcher in close contact with the participants in order to capture their perspectives on the meaning of reality—for example, interviews and photographs, and so on. Additionally, qualitative researchers study their participants in naturalistic settings while searching for the meaning and understanding of the human experience.

As with quantitative research, there is a family of qualitative research approaches that share many of the characteristics described by Denzin and Lincoln (2005). Student researchers need to become familiar with terms that summarize many of these similarities. Qualitative researchers select their participants by using **purposeful sampling**. Purposeful sampling involves the selection of participants who have key knowledge or information related to the purpose of the study. Let's say, for example, a researcher wishes to conduct a study on the experiences of parents whose preschool children attend a program that includes children with disabilities. The researcher would reach out to those parents, who are called **key informants**, who can share their experiences with an inclusion preschool.

As indicated by Denzin and Lincoln (2005), qualitative researchers use more than one method of data collection in the same study and compare the results obtained through these multiple methods. This process is known as **triangulation,** which adds thoroughness, richness, and depth of understanding to the study. Some argue that triangulation increases validity of a qualitative study. For example, a researcher collects observation data on students with disabilities who are in an inclusive classroom. Her goal is to determine the ways students with disabilities interact with regular education students. Her observations lead her to suspect that the interactions are plentiful and varied. That is the students are interacting in academic and nonacademic situations. In order to further verify her findings, she conducts one-to-one interviews with the teacher and with many of the students. These interviews allow her to confirm or disconfirm the findings based upon her observations.

Furthermore, qualitative researchers summarize their findings primarily through narrative or verbal means in order to share in-depth meaning with others. This comprehensive understanding is acquired through what researchers call **thick descriptions**. Thick descriptions involve a comprehensive description of the individual, the social context, the characteristics of the community, morals, values, and the like. Thick descriptions are written with the assistance of the **field notes**, detailed accounts of observations written by the researcher while studying the environment of the participants.

The data gathered during qualitative studies is systematically analyzed by the researcher throughout the course of the study. The method of analysis often includes **coding** the data. Coding involves the examination of the data to look for patterns, themes, or categories that emerge from the data.

While these characteristics are consistent across all qualitative approaches, there are minor differences among the designs. These are described briefly next.

Case Study While **case study** is one of the most common qualitative approaches, some case study researchers collect both quantitative and qualitative data. Although they are wide ranging in their scope and sequence, case studies typically focus on an individual, small groups, or individuals within a group and document that group's or individual's experience in a specific setting. According to Smith, as cited in Merriam (1998), case studies can be differentiated from other forms of qualitative research by the fact that these studies focus on a "single unit" or a **bounded system**. What, you might ask, is a bounded system? According to Merriam, boundedness can be determined by asking "whether there is a limit to the number of people involved who could be interviewed or a finite amount of time [for observation] . . ." (pp. 27–28).

In case studies, detailed information is gathered from multiple sources and often from the individual(s) being studied. For example, on the topic of parental involvement, a researcher could do a case study on a family or several families who are non-native-English speakers and determine how they are working with the school district and teachers to help improve their children's academic performance. Some interesting questions the researcher might think about exploring as she or he approaches the study are

- How do the parents (who are not proficient in English themselves) interact with the school in supporting and working with their children?
- Do they feel that the school is assisting them, or do they view the school as an obstacle?
- How do teachers perceive the parents' efforts to help their children?

In order to get multiple perspectives, detailed information is gathered from the parents, the children, and members of the school community. The researcher focuses on the exploration and description of the participants in the study rather than trying to generalize his or her findings to other populations (as would be the case for quantitative researchers).

In addition to gathering multiple perspectives, an important feature of case study research is the examination of multiple variables. It is recognized that the study of individuals is a complex process necessitating a study of many variables. Case study researchers embrace the belief that humans are complex creatures, and to gather comprehensive understanding many aspects of an individual's life must be investigated. Thick descriptions necessitate that researchers conducting case studies use interview, observation, documents, and artifacts as their primary tools. These data collection activities occur in the participants' naturalistic setting. In the study of non-English-speaking parents discussed previously, consider where the researcher would likely gather the data for the study. Home, school, community sites, and churches are possible naturalistic settings for this study.

In case study research, it is important to consider the length of time the researcher will spend collecting data. Remember, thick descriptions are comprehensive. Therefore, data collection often takes a long period of time. The researcher wants the participants to speak and behave naturally. In order for this to occur, the participants must feel comfortable with the researcher and develop a sense of trust and confidence. This requires time!

Ethnographic Research **Ethnographic research** is often included in the same category as case studies, and for good reason. Where case study researchers focus their energies on the interactions of individuals or small groups in specific settings, ethnographic researchers tend to investigate people in their native environments, their culture, and how their interactions in a cultural group are influenced by the larger society. Specifically, they search for meaning and "meaning is understood as being structured by culture—that is, by collectively shared and transmitted symbols, understanding and ways of being" (Miller, Hengst, & Wang as cited in Camic, Rhodes, & Yardley, 2003, p. 219).

Like cases, ethnographic research gathers information about the phenomena being investigated from multiple perspectives. However, in addition to gathering data, ethnographic researchers "filter" or assess the information gathered through the setting, recognizing that the setting itself has a role and a function in the study. Ethnographic research also requires that the researcher become familiar with the environment of the participants, to some degree, by becoming part of the group being studied. For example, an ethnographic researcher decides to examine a school within a large urban district and document how the school is trying to deal with issues of diversity. The school has been working to increase student awareness of diversity, to heighten student tolerance toward individual differences, to create a learning community, and to infuse multicultural issues into the curriculum. A researcher who clearly knows the setting, culture, and the participants gathers this information by using interviews, observations, and some document analyses. However, the researcher also recognizes that she or he has to be aware of how other outside settings or people might affect this setting (for example, the police, church members, the larger community, community leaders). Whereas the researcher is examining only one building, the larger school district, the community, and possibly the state and nation may play a role in what happens in this building.

Beginning educational research students often confuse case study and ethnographic research because of their similarities. One way to differentiate the two approaches is by examining the scope of the study. Ethnographies tend to be of larger scope and have greater breadth than case studies. They require that the researcher spend extensive time gathering data in the field (environment) and in sustained community contact. By engaging in these practices, ethnographers can gain meaning of the culture and social systems being investigated.

Phenomenological Research Like case studies, **phenomenological research** is a common qualitative approach. Phenomenological researchers work to describe lived experiences. By doing so, they attempt to capture the "essence" of the human experience by describing with great precision the personal experiences of the participants of the study.

Like other qualitative researchers, phenomenologists are interested in understanding and recording the social and psychological perspectives of the participants in the study. Phenomenological researchers attempt to capture everyday experiences of those being studied. Moreover, they stress the importance of capturing the view of reality described in the words of the participants. According to van Manes (1990), phenomenology is the "description of the lived-through quality of lived experience and the description of meaning of the expressions of lived experiences" (p. 25).

In order to capture the lived experiences of the participants, the researcher relies on an honest and detailed account of the life experience being investigated. As you might imagine, this is not an easy task. The researcher asks participants to reveal their experiences by providing detailed information about those experiences. This requires that the participant and the researcher make important psychological connections. The participant must develop a sense of trust in the researcher which often requires that the researcher consider the participant as a co-researcher. In this way the participant becomes part of the study and controls, to some degree, the type of questions that are asked and the way they are asked. The researcher must also be keenly aware of her or his own perspectives and personal biases. Remember, the goal is to capture and describe the experiences of the participants, and this means that the researcher must keep her or his personal perspectives out of the discussion. Some phenomenological researchers write their biases about the phenomena on a piece of paper, tear up the paper, and "symbolically" remove those perspectives from the study.

The researcher gathers information about the personal experiences of the participants primarily through phenomenological interviews. Phenomenological interviews are made up of open-ended questions through which the participants can fully share their personal experiences. Once the interviews are complete, they must be transcribed precisely, since they become the raw data for the study. The transcribed data is then read completely to gather a holistic picture or an understanding of the story being told from beginning to end. The researcher looks back through the transcripts, focusing on how to describe the meaning of the experience. For an example of a published study that used phenomenological methods to study the experiences of women in dance therapy, see Mills and Daniluk (2002).

Consider the following example: Take a look at the person sitting next to you in class. You both are sitting in the same course, at the same college, with the same professor; yet, the way you perceive the reality of this graduate experience is quite different. You each bring a history of personal experiences, attitudes, behaviors, and emotions, all of which influence how you view this shared experience. The phenomenologist role is to "give voice" to those personal perspectives.

Narrative Research **Narrative research** in education continues to grow in popularity because of the richness of data it produces. In narrative research, researchers portray the lives of people in a particular setting or context through story telling. According to Connelly and Clandinin (1990), "Narrative method, in its simplest terms is the description and **restorying** of a variety of educational experiences. A researcher's narrative account of an educational event may constitute a restorying of that event and to that extent is on a continuum with the process of reflective restorying that goes on, one way or another, in each of our educational lives" (p. 4). In their definition, Connelly and Clandinin touch upon several important features of narrative research in education. First, narrative research focuses primarily on describing the experiences of an individual. These experiences are described through a collection of stories that report the individual's experiences and by discussing the meaning of those experiences for the individual. Second, these experiences are restoried through the researcher. Restorying involves the researcher gathering the stories, analyzing the stories, providing causal links among thoughts, and retelling the stories in a chronological sequence. This process, of course, involves close contact and interaction with the person or persons telling the story.

Like other forms of qualitative research, the process of narrative research involves the identification of the phenomena or question to be studied. Let's say that a researcher is interested in studying the experiences of teenaged fathers who continue to go to school. In order to conduct a narrative study, the researcher would purposefully select an individual (in this case a teenaged father) to learn about his experiences. First and foremost, the researcher must select an individual who is willing to tell his story. Once the participant is identified, the researcher conducts a series of interviews to allow, in this case, the teenaged father an opportunity to tell his story. Interview questions often help the researcher to probe more deeply the phenomena being studied. Of course, the interview questions are only a framework, and the participant may tell his story beyond the questions being asked. As you might imagine, the researcher needs to develop a bond of trust with the participant in order to obtain the best possible story. All stories are recorded by the researcher for later transcription and retranscription (note that most narrative researchers transcribe the stories more than once to be certain accurate information is obtained).

Some narrative research uses multiple types of data in order to provide a complete picture of the story. In one type of narrative research, biography, the researcher might interview many individuals who knew the subject of the biography and also review documents related to his/her lives such as letters, diaries, newspaper articles or web pages.

The researcher then examines all data for patterns or codes to identify the setting, main characters, problems, and potential solutions. The restorying involves the organization of the story into a logical sequence with a beginning, middle and end while bringing the voice of the participant to the forefront.

Unlike ethnography, narrative research focuses on telling a micro- rather than a macro-story. This means that an important key element for narrative research is the story of the individual and his or her experiences in a particular context.

There are many different forms of narrative research, and the processes used in these different approaches will vary slightly. Much more detail on narrative research is presented in chapter 6.

Research Approaches Using Qualitative or Quantitative Approaches (or Both)

Several approaches to research are more flexible and may use either qualitative or quantitative approaches or use both in a single study. This book focuses on two approaches most relevant to the practitioner-researcher: action research and program evaluation. For discussion of more complex mixed-method approaches, see Creswell (2005).

Action Research **Action research** is designed to enhance and improve current practice within a specific classroom, school, or district. Typically, it is a type of research undertaken by practitioners who have identified problems they wish to solve or who would like to find ways to enhance their own teaching or student learning, or both. Action research is a cyclical process that begins with the researcher or practitioner engaging in reflection on his or her practice. Specific and practical questions are asked. Are all students learning using this instructional approach? Are all clients in the after-school group therapy sessions participating with a certain type of question? As part of this reflection process, often both qualitative and quantitative data are collected on the study participants. This reflection may lead the researcher to decide that systematic change is warranted. The next step takes the researcher to the literature and/or colleagues to investigate possible ways or techniques to improve practice. Once an action plan is decided upon, the researcher applies the technique and then measures its impact. The study continues with ongoing reflection to determine whether any further modifications in practice are needed.

It is entirely possible that the study will lead to the conclusion that the strategy being used is in fact working. For example, Pasko (2004) studied her own third grade classroom to see how the students connected mathematical concepts to literature that they read independently. Her results provided support that her interdisciplinary approach to teaching was working for her students and suggested new ideas that she might try to improve her approach.

Action research generally includes a three-step process:

1. Identification of the problem(s) through careful observation and reflection
2. Planning and taking appropriate action (the study)
3. Using the findings to determine if teaching and learning have improved or if further changes are needed

This type of research continues to grow in use because educational practitioners find it an empowering and collaborative activity.

Program Evaluation The field of education is filled with programs designed to improve both learning and teaching. Examples of these programs include a reading intervention program designed to help struggling readers or a teacher-training program designed to help teachers integrate technology into lessons. **Program evaluation** is designed to attempt to determine the level of success or failure of such educational programs and to make decisions about such programs. Although program evaluation uses quantitative and qualitative methods, its overall purpose is different from most other types of research. Whereas quantitative and qualitative researchers study programs, findings from such studies typically are slow to change or improve the programs themselves. In program evaluation, however, findings are often used for ongoing or short-term decision-making purposes, and programs can be changed or "improved" based on the results of a single evaluation. In some extreme cases, a program might even be eliminated based on such evidence. Most program evaluation approaches use two types of feedback loops for reporting findings: formative feedback and summative feedback.

Formative data are collected and provided to program developers as the program is occurring, with the hope that such evidence will support the needed changes. For example, if one is evaluating a new reading program and the instruction is not being delivered according to the program's goals, the evaluator would provide this information to the program director so that the instruction could be improved. Although some quantitative researchers use formative feedback loops, it is the potential for action to be taken on the feedback that makes program evaluation distinct from quantitative approaches. For example, experimental or quasi-experimental researchers would not dream of altering the program or treatment (the independent variable) as it was being studied. After all, if the study showed an increase in student performance, to what could the results be attributed? The program before the improvements? The program after the improvements? A combination of the two? In addition to collecting and providing formative feedback, program evaluation researchers attempt to collect summative data.

Summative data focus on determining whether a program's goals were met. Examples of summative data are changes in students' reading scores, number of people served by the program, and job satisfaction ratings. Program evaluators tend to use both formative and summative information in identifying areas in need of improvement and in determining a program's success or failure.

Note that evaluation studies are often published in specialized journals such as the *American Journal of Evaluation*. They also appear as reports provided directly to the client who hired the evaluator.

SUMMARY

This chapter provides you with executive summaries of each research approach and important terms or key concepts associated with each approach. The summaries offer an overview of the purpose, the types of questions addressed, and the methods used in each approach. It is our belief that this will provide you with the framework needed to read research in your fields. As you read research, it is helpful to think about which approach the researcher is utilizing and why. You should be better able to determine whether the research approach selected is effective at answering the research question. Furthermore, this knowledge will prepare you to more effectively read and understand the later chapters in the textbook, where we provide in depth discussion of each approach.

The chapter delineates the different approaches used under the broad categories of qualitative and quantitative research. While it is important to know the types of research associated with qualitative and quantitative research, it should be noted that many researchers combine more than one approach when conducting research. In fact, more and more studies use both qualitative and quantitative data collection tools in the same study. This is especially the case with research conducted by the practitioner because of easy access to different types of data. For example, a teacher who conducts an action research study can gather test scores from her students and interview them on the new instructional technique she is using in math. Researchers may also collect both qualitative and quantitative data to verify the data they collect. For example, if a researcher collects observational data on students in a classroom with an observational checklist (a quantitative measure), she may also decide to interview the teacher (qualitative measure) to validate the results of her observation.

KEY CONCEPTS

action research	correlational research	field notes
baseline phase	correlation coefficient	formative data
bounded system	demographic items	hypothesis
case study	dependent variable	independent variable
categorical variable	effect size	key informants
coding	ethnographic research	narrative research
constructs	experimental research	negative correlation
continuous variable	ex post facto research	nominal variable
control group	extraneous variables	nonexperimental research

operational definition
phenomenological research
pilot study
population
positive correlation
program evaluation
purposeful sampling
quasi-experimental research
random assignment
random selection
reliability
response rates
restorying
sample
sampling bias
subquestions (in descriptive and qualitative research)
summative data
thick descriptions
threats to internal and external validity
treatment phase
triangulation
validity
variable

DISCUSSION QUESTIONS OR ACTIVITIES

1. Pick two research approaches discussed in this chapter and write research questions that could be addressed using each approach.
2. Identify a research topic and discuss how it might be studied using both a quantitative and qualitative approach.

SUGGESTED READINGS

Mills, L. J., & Daniluk, J. C. (2002). Her body speaks: The experience of dance therapy for women survivors of child sexual abuse. *Journal of Counseling and Development, 80*(1), 77–85.
van Manen, M. (1984). Practicing phenomenological writing. *Phenomenology & Pedagogy, 2*(1), 36–69.

CHAPTER THREE

DESCRIPTIVE STATISTICS

CHAPTER OBJECTIVES

- Differentiate among frequency distributions, frequency polygons, and histograms and interpret data produced by each of these
- Describe the appropriate situations when a researcher would use a frequency polygon versus a histogram
- Describe the characteristics of a normal distribution
- Describe how frequency scores are distributed and discuss how frequency scores are clustered for a distribution that is positively skewed, negatively skewed, and bimodal
- Identify and give examples for the three types of measures of central tendency
- Know how to calculate and interpret measures of central tendency
- Identify and give examples of the different types of measures of variability
- Know how to calculate and interpret variance and standard deviation
- Define a correlation coefficient and give an example of one that shows a strong positive relationship between two variables and one that shows a strong negative relationship

CHARACTERISTICS OF DATA

In the previous chapter, brief summaries were presented of the different approaches used by quantitative researchers. A commonality among all approaches is that numerical data is gathered and analyzed. **Quantitative data** is numerical information (for example, attendance rates, standardized test scores, rank in class). Numerical information is used by all educational professionals to measure educational outcomes as well as to summarize and draw conclusions about quantitative research.

It would be nice if measuring educational outcomes was as easy as stepping onto a scale. However, the process is somewhat more complicated than that! Just as you would need to know whether your scale is based on pounds or kilograms, researchers and practitioners need to know what type of scale of measurement is being used in the test or assessment tool. You may ask why. The scale of measurement (or type of data) will dictate the type of statistics that can be used to interpret the data. The scale for an educational measure is categorized based on the range of values or scores produced by the measuring instrument. Specifically, measuring instruments produce four levels or scales of measurement: nominal, ordinal, interval, and ratio scales (Table 3.1).

Nominal scales measure variables that are categorical: variables that represent discretely separate groups or categories that are not ordered. This is considered to be the lowest level or scale of measurement. Each person or school is placed into a single category (for example, category A, category B) based on its score. For example, let us say that a district reported the number of students who qualified for free or reduced-cost lunch—a measure often used to estimate the number of students from low-income families. Students would be categorized

TABLE 3.1 Scales of Measurement

Scale of Measurement	Description	Examples
Nominal	Categories	Private versus public school, parenting style
Ordinal	Categories, plus ranking	Class rank, Olympic medals (gold, silver, bronze)
Interval	Categories, ranking, equal spacing	Reading test scores, grade point average
Ratio	Categories, ranking, equal spacing, plus true zero	Age, weight, and time

by assigning them to either the Qualifies for Free or Reduced Lunch category or Does Not Qualify for Free or Reduced Lunch category. Other variables that might be measured using a nominal scale would be gender, ethnicity, presence or absence of a disability, or type of leadership style. The data would then be summarized numerically for each category. Limited statistics can be used to analyze data produced by nominal assessments. More than likely, the number of persons or schools in each category would be reported.

Ordinal scales, like nominal scales, categorize persons or things. However, they also place the categories into rank order from highest to lowest or from lowest to highest. Therefore, they allow us to determine who did the best or who did the worst. A commonly used example of ordinal data is high school class rank. Consider the following as four high school ranks and associated grade point averages (GPAs):

Rank	*GPA*
1	4.0
2	3.8
3	3.6
4	3.2

Students are placed in order based on their GPAs, but the data are entered as rank in class. Notice that even though you can now determine the highest to lowest high school ranks, the distance between the scores or the intervals is not equal. The distance between rank 1 and 2 is 0.2 GPA points, and the distance between rank 3 and 4 is 0.4 GPA points. This limits the type of statistical tests that can be applied to the data and also limits the conclusions one can draw about differences between students at different ranks. As with the analysis of nominal data, the type of statistics that can be used is very limited. Typically, the number of persons in each rank is reported. GPA is an interval scale measure, discussed next.

Interval scales categorize and rank order the data, thereby including the characteristics of nominal and ordinal data. However, with interval data it is assumed that the distance between each score is equal. For example, Scott's midterm exam grade in his educational research course is a 65, whereas his roommate Roberto scores a 69. How many points separate Scott and Roberto? No, this is not a trick question! Four points separate the two students. In the same class, Josie scores a 95 and Maria scores a 99. How many points separate these two students? Right, four points. It is assumed that the difference between Scott and Roberto is the same as the difference between Josie and Maria. Although interval data are more sophisticated and precise than either nominal or ordinal data, there are still certain limitations. First, in interval data, there is no true zero point. What does this mean? Let us begin by defining the word *zero*. If you were asked for a definition of zero, you would likely say something like "There is nothing" or "There is an absence of a quality." This is true. However, if poor Harold got a zero on the midterm, does this mean he has no knowledge at all of the topics covered in the class? Probably not! He certainly does not know much about educational research, but he probably has some content knowledge. So in this case, to say that a score of zero means the absence of something is not totally accurate. In addition, interval data have an arbitrary maximum point. Let us say that Helena gets a score of 100 out of 100, and Ramone gets 50 out of 100 points on the midterm. Does this mean that Helena possesses all the possible knowledge of educational research and that Ramone possess 50% of the content? Certainly not! The exam simply went up to 100 points. Receiving a score of 100 does not indicate total mastery of the content, and receiving a score of 50 does not mean mastery of half of the content. Because there is no true zero point, Helena does not have 50% more knowledge about educational research than Ramone. Most standardized educational measures use interval scales. With interval data, the range of statistical techniques that may be used to analyze the data is quite large. Measures of central tendency, variability, and relationship (see following) as well as inferential techniques (discussed in chapter 11) are used in the analysis of these types of data.

Ratio scales are often considered the types of measurement that produce the most precise data. Ratio scales include the properties of nominal, ordinal, and interval variables and also include a true zero point. Consider height: a child who is 3 feet 1 inch tall is 50% of the height of a person who is 6 feet 2 inches. This is true because the variable height has a true zero point. A height of zero literally means the absence of height. Also, there is not an arbitrary limit to how tall someone could be. (Professional basketball players get taller each year!) Ratio variables are seen often in the physical sciences but rarely in education. As described for interval data, ratio data can be analyzed using many different statistical techniques.

SUMMARIZING DATA USING DESCRIPTIVE STATISTICS

Once the scale of measurement has been determined, **descriptive statistics** can be used to summarize data using either graphical or mathematical procedures. Almost every study using a quantitative measure will use descriptive statistics to depict the patterns in the data. Similarly, in educational practice, we often want ways to describe the overall performance of students, teachers, administrators, or schools on some measure. Descriptive statistics are one tool for summarizing this performance for both research purposes and, increasingly, for assessment of educational outcomes at the classroom, school, and district levels. Practitioners are increasingly asked to use statistics to monitor student learning and interpret student performance on classroom and standardized measures.

Frequency Distributions

One of the first questions educators ask when a group has been measured is, "How did we do?" To answer this question, one needs a way to see how the group as a whole did as well as individual students. One way to depict the overall performance of a group is to display the frequency of each score in a **frequency distribution**. The word *distribution* is used to describe the range of scores and their overall frequencies. A frequency distribution displays each score and the frequency with which it occurred in either a table (called a **frequency table**) or in a graph. The scores are ordered from highest to lowest, and the number of persons obtaining each score is listed as the frequency in the second column. For example, in Table 3.2, the number of books that students read (scores) is reported in the first column, and the frequency or number of students who read that many books is reported in the second column.

Sometimes possible score values are grouped rather than being listed individually to make the patterns clearer, as in Table 3.3.

Scores may also be displayed in a graph with the scores listed along the x or horizontal axis and the frequencies listed along the y or vertical axis. If the data represent an ordinal scale of measurement or higher, they are usually connected by a line, as in Figure 3.1, and the graph is referred to as a **frequency polygon**. If the data are categorical, each category is represented by a separate bar, as in Figure 3.2, and the graph is called a **histogram**.

So what does a frequency distribution tell us about the scores? At a glance, it shows us how the scores distribute. This means how closely bunched together or how spread out the scores are, which scores are most frequent and which scores are least frequent, and whether there are **outlier scores**. Outlier scores are scores that are very different from the rest.

Distributions of educational measures can take several different forms. For example, when large groups are randomly sampled and measured, the distribution often has a shape that is called a **normal distribution**. If the distribution is approximately "normal," it will look bell shaped and symmetrical, with the highest point on the curve or most frequent scores clustered in the middle of the distribution. Frequencies of scores will decrease as the distribution moves away from the middle toward either the high or low end of the score values. Most

TABLE 3.2 Frequency Table of Number of Books Read by 30 Students

Number of Books	Number of Students
20	1
19	0
18	0
17	0
16	1
15	1
14	0
13	1
12	1
11	1
10	1
9	2
8	1
7	2
6	4
5	5
4	4
3	2
2	2
1	1

TABLE 3.3 Sample Grouped Frequency Table

Number of Books	Number of Students
17–20	1
13–16	3
9–12	5
5–8	11
1–4	10

FIGURE 3.1 Sample Frequency Polygon

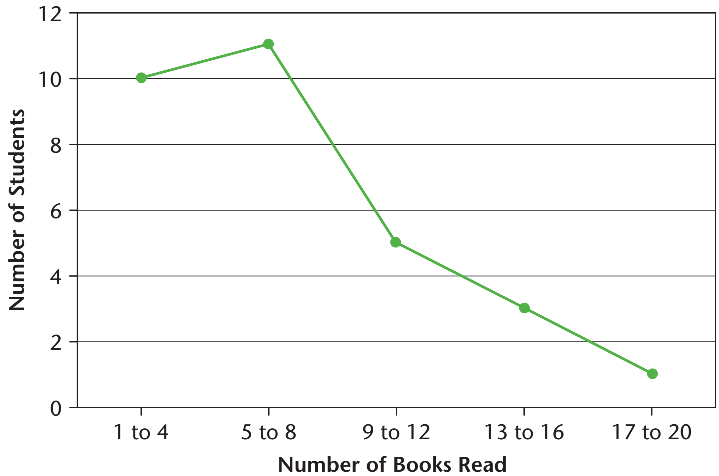
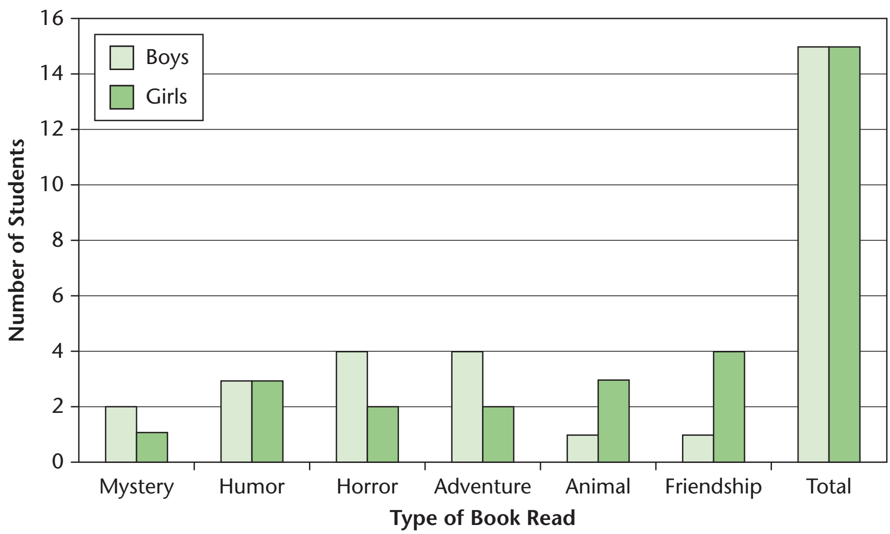


FIGURE 3.2 Sample Histogram



standardized tests present data from a norm group that shows a normal distribution of scores. The symmetrical nature of this distribution allows one to examine a wide range of differences among scores through the full range of the scores. Skewed distributions are asymmetrical, meaning the scores are distributed differently at the two ends of the distribution. In a **negatively skewed distribution**, most of the scores are high, but there are a small number of scores that are low. In a **positively skewed distribution**, most of the scores are low, but there are a few high scores. Figure 3.3 shows the test scores for a class in educational research under two different skewed conditions. In which class would you rather be a member?

The outliers in a skewed distribution pull the tail of the distribution out in that direction. So a negatively skewed distribution has a low flat tail at the low end of the scores, and a positively skewed distribution has a low flat tail at the high end of the scores.

FIGURE 3.3 Examples of Positively Skewed and Negatively Skewed Distributions

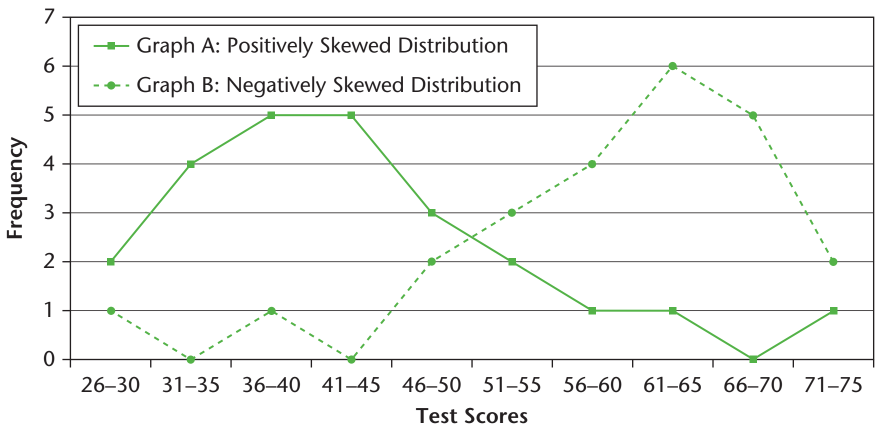
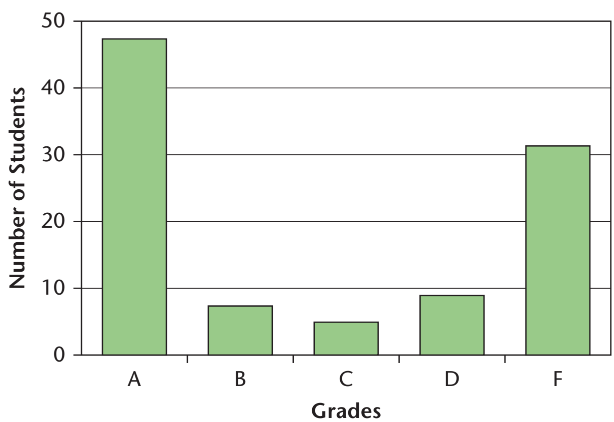


FIGURE 3.4 Grades in Distance Education Classes



Some distributions, called **bimodal distributions**, have two clusters of frequent scores seen as the two humps in the distribution, as in Figure 3.4. Notice in the graph that the number of students earning As or Fs far exceeds students getting Bs, Cs, and Ds. Although the number of As and Fs are not equal, both are much higher than the other categories. Bimodal means that the distribution has two modes, a measure discussed following.

Measures of Central Tendency

Although frequency distributions show the patterns in scores, it is useful to summarize the performance of a group using a single score for the typical or average performance of a group. These are what researchers call **measures of central tendency**, and the three most common are the mode, mean, and median.

Mode The **mode** is the score in a distribution that occurs most frequently. In a frequency polygon, the mode is the score represented by the highest point on the curve. As you saw in Figure 3.4, some distributions can have more than one mode if two or more scores have similarly high frequencies. (That is why the distribution in Figure 3.4 is called a *bimodal distribution*.) The mode is a somewhat imprecise measure of central tendency because it simply summarizes the frequency of a single score. If the distribution is asymmetrical, the mode may not be a precise estimate of central tendency.

Mean The **mean** is the arithmetical average of a set of scores. To find the mean, simply add up the scores and divide by the number of scores. If scores or numbers are in a frequency table, first multiply each score by its frequency. Table 3.4 presents the number of pages printed at campus computer terminals by one class of students during a single semester and illustrates the calculation of the mean.

A college recently considered using the mean number of printed pages to set a new policy for the number of free pages students would be allowed to print. They reasoned that because the mean represented the average number of pages printed, the new policy would allow the average student to do all or most of his or her printing for free. The mean is a widely used measure of central tendency in research and in educational practice.

In a skewed distribution, however, the mean can be misleading as a measure of central tendency. The extreme scores in a positively skewed distribution will result in the mean being higher than the central tendency of the group, and the extreme scores in a negatively skewed distribution will pull the mean down toward the low end. Again, using the printing example, the mean for all of the students at that college was much higher than the average for just one class because a small

TABLE 3.4 Calculation of Mean from a Frequency Table

Number of Pages Printed	Number of Students	Number of Pages × Number of Students
300	1	300
280	3	840
260	5	1,300
240	6	1,440
220	4	880
200	2	400
180	2	360
160	1	160
	Total = 24	Total = 5,680

Note: The mean is calculated as: 5,680 ÷ 24 = 236.67.

number of students printed more than 1,000 pages a semester, and one even printed a million pages! (Of course, the new policy was intended to limit the printing of students like this who were abusing the system.) However, the mean estimated the average number of printed pages per student at a level much higher than was typical for most students.

Income is another measure for which the mean is likely to be a poor estimate of central tendency. Income distributions are typically positively skewed. Most of us earn middle to low incomes, but there are a small number of people who earn very high incomes. Imagine what would happen to the mean income of your community if one of the residents was Bill Gates! In a skewed distribution, the best measure of central tendency is the median (discussed later in the chapter).

The symbol used to represent the mean is a capital X with a bar on top ($\bar{X}$), while X standing alone represents the raw (or actual) score. What makes the mean such an important measure of central tendency is the fact that the mean reflects every score in the distribution. The fact that the mean is responsive to all of the scores in the distribution leads to an important mathematical principle. If the deviations of each score ($X - \bar{X}$) from the mean are summed, the resulting value will be equal to zero. If this sounds confusing, examine the following data set of possible test scores (let's call it distribution Z):

40
50
60
70
80
90
100

The sum of (Σ) the scores is equal to 490 and the $\bar{X}$ is equal to 70. A **deviation score** ($X-\bar{X}$) is an indication of the distance of the scores from the mean. If the mean is subtracted from each score on the low end of the distribution, the sum of the deviations is equal to −60 (see calculations following).

$$X-\bar{X}$$
$$40 - 70 = -30$$
$$50 - 70 = -20$$
$$60 - 70 = -10$$

If the mean is subtracted from each score on the high end of the distribution, the sum of the deviations is equal to +60 (see calculations following).

$$X-\bar{X}$$
$$80 - 70 = +10$$
$$90 - 70 = +20$$
$$100 - 70 = +30$$

Therefore, the total of all of the deviations (adding +60 and −60) equals zero. If the deviation scores are examined, the researcher can get a sense of the variability (discussed in greater detail following) of the distribution. Remember, the deviation scores represent how far each score is from the mean. Larger deviations scores are an indication of greater variability in the distribution. For example, let's say that a second distribution (we will call it distribution Y) produces the following set of scores:

20
20
25
80
85
90
100

In this distribution, the sum (Σ) of the scores equals 420 and the mean is equal to 60. The deviation scores for this distribution are as follows. On the low end of the distribution the sum of the deviation scores equals −115 (see calculations following).

$$X-\bar{X}$$
$$20 - 60 = -40$$
$$20 - 60 = -40$$
$$25 - 60 = -35$$

The sum of the deviation scores on the high end of the distribution equals +115 (see calculations following).

$$X - \bar{X}$$
$$80 - 60 = +20$$
$$85 - 60 = +25$$
$$90 - 60 = +30$$
$$100 - 60 = +40$$

The summed total deviation scores (from the high and low ends of the distribution) correctly equals zero ($-115 + 115 = 0$).

Where is the greater variability? You are right if you said distribution Y! There is certainly less clustering around the mean in distribution Y and therefore more variability. This concept of deviation scores becomes more significant when we calculate variance and standard deviation later in the chapter.

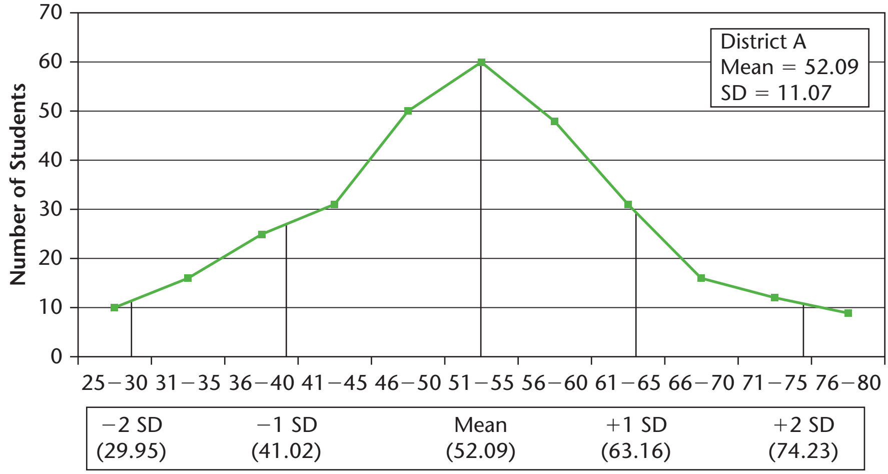
Median The **median** is the score that divides a distribution exactly in half when scores are arranged from the highest to the lowest. It is the midpoint of distribution or the score in the distribution at which half of the scores are lower and half are higher. To find the median, you would divide the number of scores in a distribution by 2 and round up if there is a decimal. Then count down from the top of the distribution until you have counted that number of scores. If a distribution has an even number of scores, the median is the score halfway between the two scores in the middle. If a distribution has an odd number of scores, the median is an actual score. Consider the following simple example. If a distribution includes the scores 60, 70, 80, 90, 100, what is the median? If you said 80, you are correct. This distribution has an odd number of scores (five), so the median is an actual score. If the distribution only includes scores of 60, 70, 80, 90, what is the median? If you said 75 you are correct. In this case, there are four scores (an even number), and the median is the midpoint between the two middle scores.

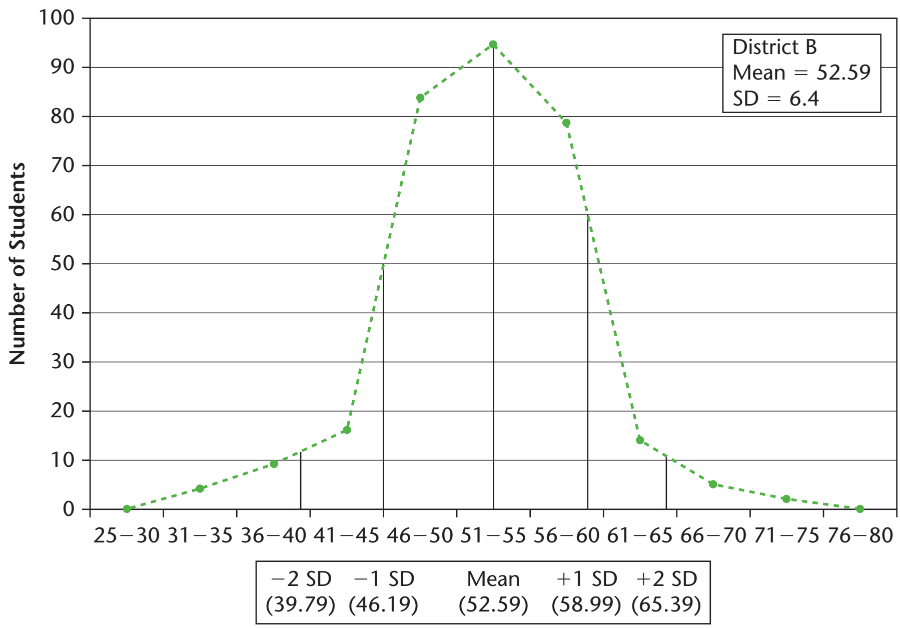
The median is a stable measure of the central tendency of a set of scores. This means that it is not affected much if there are a few outlier scores in a distribution that are much different from the rest. To return to our example of income distributions, if Bill Gates moved into your community, the distribution of income would not change greatly because the addition of one person would simply move the median up one number in the distribution.

Measures of Variability

So now you know several statistics that can be used to describe the average or typical performance of a group. However, two groups might have the same mean even if they differ considerably in the spread of the scores in their distributions. Consider the two distributions of test scores in Figure 3.5.

FIGURE 3.5 Distributions of Test Scores for Two School Districts





Both districts have approximately the same average performance in their students' test scores; the mean for district A is 52.09, and the mean for district B is 52.59. However, district A's scores are much more variable than district B's. Descriptive statistics can provide a way to summarize the amount of variability in the scores in a distribution, that is, how widely spread out or scattered the scores are. Why is variability of interest to researchers and even practitioners? Essentially, variability is one of the fundamental reasons to conduct research. To what extent does an instructional strategy or a behavioral treatment work for some children and not others? These important questions can be answered by looking at the variability in scores used to measure the efficacy of the treatment. In order to reinforce your understanding, let's examine two different examples, one from a teacher's perspective and one from a school psychologist's perspective.

Mr. Karne is a 12th-grade teacher covering a unit on the Maya civilization. He has the students read the textbook and read reports by expert archeologists. After these activities, he conducts an in-class discussion about the reasons for the collapse of the ancient Mayas. Mr. Karne gives an exam (which closely matches the content covered) and finds that some students perform well and others get very low scores. In other words, there is variability in the scores. This variability or individual differences in student performance is of interest to Mr. Karne because he hoped (as do all teachers) that all students would be successful. Why is it that the same instructional strategy resulted in positive outcomes for some students, while other students continued to struggle with the content? The fact that there was variability in students' performance will require that Mr. Karne to further examine the instructional technique itself and the way it was implemented in his classroom. Mr. Karne's goal is to decrease variability in student scores because he wants all students to pass.

Ms. Gomez is a school psychologist leading an after-school group therapy program for girls with eating disorders. She administers a self-esteem inventory at the beginning of the program and determines that all of the girls could benefit from a workshop designed to enhance their self-esteem. The literature reveals that this workshop is highly successful at improving the self-esteem of girls with eating disorders. She plans and administers a comprehensive self-esteem workshop. Of the 20 girls in the program, 12 girls improve and 8 girls show no change in their self-esteem. She recognizes that there is variability in the way the girls react to the treatment. Why? What is it about the girls or the treatment that caused the variability? Once again, the fact that variability exists would encourage Ms. Gomez to further study her self-esteem program and the way it was implemented.

We hope that these two examples help to explain why researchers and practitioners are interested in variability. Following are three examples of the ways that variability is measured.

Range

The **range** is the difference between the highest and lowest scores in a distribution. To find the range, you subtract the lowest score from the highest score. This indicates how many points separate these two scores. For example, let's say that Mr. Karne gives a 60-point exam on the Maya civilization. If the highest score on Mr. Karne's test was a 58 and the lowest score was 18, the range would be 40. Although it is easy to calculate, the range is not a precise or stable measure of variability because it can be affected by a change in just one score. For that reason, most educational researchers use variance and standard deviation as measures of variability.

Variance Remember those deviation scores we discussed earlier? We can almost hear the resounding sounds of your collective sighs! As discussed previously, summing deviation scores results in zero and tells us nothing about the variability of a distribution. We could compare one distribution to another, which would give us a sense of a distribution's variability (as we did when comparing distribution Z to distribution Y). But what if you wanted to examine only a single distribution? In order to determine the variability of a distribution with a single measure, the **variance** is calculated. The variance is the average or mean of the squared deviation scores! Variance is denoted by the symbol s^2. The formula is

$$s^2 = \frac{\sum (X - \overline{X})^2}{n - 1}$$

Note that the numerator of the formula is the sum of the squared deviations of each score from the mean $\Sigma(X - \overline{X})^2$. (You might remember from our discussion of the mean that the sum of the deviations from the mean is always zero. This necessitates the squaring of each deviation.) The sum of the squared deviations is often referred to as ss or **sum of squares** or **the sum of the squared deviations from the mean**. The simplified formula for variance, using ss, is

$$s^2 = \frac{ss}{n - 1}$$

Using distribution Z described above, let's calculate the variance.

Calculating the Variance for distribution Z Scores: 40 50 60 70 80 90 100

Step 1. Calculate the mean of the distribution. Add all of the scores and divide by the total number of scores. The mean of distribution Z equals 70.

Step 2. Subtract the mean from each score. This will give you the deviation scores.

X (the scores)	−	$\overline{X}$ (the mean)	$X-\overline{X}$ (the deviation score)
40		70	−30
50		70	−20
60		70	−10
70		70	0
80		70	+10
90		70	+20
100		70	+30

If we summed the deviation scores, what would we get? If you said zero you are right!

Step 3. Square each deviation score.

$X-\overline{X}$ (the deviation score)	$(X-\overline{X})^2$ (squared deviations)
−30	$(-30) \times (-30) = 900$
−20	$(-20) \times (-20) = 400$
−10	$(-10) \times (-10) = 100$
0	$(0) \times (0) = 0$
+10	$(+10) \times (+10) = 100$
+20	$(+20) \times (+20) = 400$
+30	$(+30) \times (+30) = 900$

Step 4. Sum the squared deviations.

$(X-\overline{X})^2$ (squared deviations)

900
400
100
0
100
400
900

$\Sigma(X - X-^2 = 2800$; this is sum of squares or ss)

Step 5. Determine *n* (the total number of scores in the distribution). In this example n = 7.

Step 6. Insert numbers into variance formula.

$$s^2 = \frac{\sum(X - \overline{X})^2}{n-1}$$

2800/6 = 466.7

Now test yourself using the steps described and calculate the variance for distribution Y. The scores are 20, 20, 25, 80, 85, 90, and 100. The sum of the squared deviations is equal to 7950. Divide 7950 by 6 (n − 1). The variance is equal to 1325.

You might be asking what the variance tells you in a practical sense. What does it mean? Variance is used mostly for inferential statistics (covered in chapter 11) and is of little interpretive use as a descriptive statistic. The solution to this is to un-square the variance (which is what happens when we look at the next concept of standard deviation).

Standard Deviation The **standard deviation** is the square root of the variance and is the average distance between each of the scores in a distribution and the mean. Now before your eyes glaze over or you begin to wonder how researchers can come up with so many meaningless concepts, let us consider why you would want to calculate such a statistic. Remember that the goal is to summarize the average amount of variability in a set of scores. We know that the mean represents the average performance of a group or the center of the group. So one way to describe variability is to consider on average how far each score is from that center score. It is called the standard deviation because it represents the average amount by which the scores deviate from a mean. While many hand-held calculators can easily produce a standard deviation, we believe that examination of the formula and hand calculation of the formula furthers students' comprehension of the conceptual aspects of standard deviation.

To calculate the formula for standard deviation using the variance calculated earlier for distribution Z (see above), all you have to do is take the square root of 466.7 (the variance of distribution Z). The square root of 466.7 is 21.6

$$s^2 = \sqrt{\frac{\sum(X - \overline{X})^2}{n-1}}$$

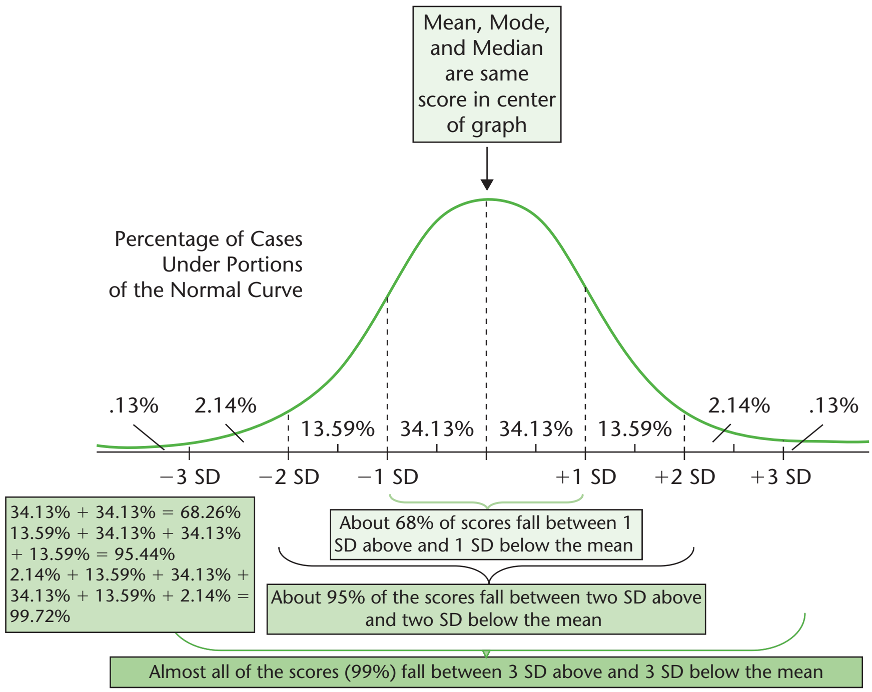
and therefore the standard deviation of distribution Z. What does this mean? The standard deviation gives you a sense of the average dispersion that exists in the distribution. The smaller the standard deviation, the less dispersion and the closer the scores are to the mean. Standard deviation also allows us to examine a particular score and determine, in standard units, how far the score is from the mean of the distribution. For example, in distribution Z, the standard deviation is 21.6 and the mean is 70. The student scoring 100 has scored greater than one standard deviation *above* the mean. The student scoring 40 has scored greater than one standard deviation *below* the mean.

Now calculate the standard deviation of distribution Y. Remember from above the variance is 1325. What is the standard deviation? If you said about 36.4 you are right! We've now proved our earlier statement that there is more variability in distribution Y than in distribution Z.

As you can see, the calculation of standard deviation is relatively simple, and most research studies report the standard deviations for the data collected. For any distribution of scores, the standard deviation will be a specific number, typically including decimals because its calculation rarely results in a number. The larger the number, the more variable are the scores in the distribution. In Figure 3.5, the standard deviation for the scores in district A is 11.07 and is 6.4 for district B. If you examine the scores, you will see that, on average, most of them lie within 2 standard deviations of the mean for both districts. That is, if you subtracted each score from the mean, the resulting number would be a value that was less than 2 times the standard deviation (22.14 for district A and 12.8 for district B). The standard deviation is widely used in part because of its relationship to the normal distribution, visually represented as the normal curve (displayed in Figure 3.6), which forms the basis for comparing individual scores with those of a larger group.

Normal Curve The **normal curve** is a distribution that has some unique and useful mathematical properties. One useful feature of the normal curve is that a certain percentage of scores always falls between the mean and certain distances above and below the mean. These distances are described as how many standard deviations above or below the mean a score falls. Approximately 34% of the scores fall between the mean and one standard deviation above it. Similarly, approximately 34% of the scores fall between the mean and one standard deviation below it. So about two-thirds (more precisely, 68%) of the scores will be in the area between one standard deviation above and one standard deviation below the mean. Another 13.5% of the scores will fall in the area that is more than one standard deviation above the mean but less than two standard deviations above the mean. So between the mean and two standard deviations are 47.5% of

FIGURE 3.6 Normal Curve



the scores (34% plus 13.5%). Finally, another 2.5% of the scores are higher than two standard deviations above the mean. Because the curve is symmetrical, 13.4% of the scores will also be in the area that is between one standard deviation below the mean and two standard deviations below the mean. Also, 2.5% of the scores will fall below two standard deviations below the mean. Figure 3.6 illustrates the percentage of scores in each part of the normal curve. These percentages are useful in comparing the performances of persons who have taken an educational measurement.

Interpreting Measures of Central Tendency and Variability

Researchers often use tables and graphs to aid in the interpretation of both measures of central tendency and variability. These tables summarize the performance of the groups and provide a quick way to get an overall picture of the groups' performance. Table 3.5 shows the means and standard deviations for a study examining the effectiveness of teaching self-regulated learning strategies to students

TABLE 3.5 Means and Standard Deviations for the Number of Words Written

	Baseline	Baseline	Baseline	Treatment	Treatment	Treatment
Mean	95	95	97	106	115	142
Standard Deviation	8	11	8	10	9	13

Source: Adapted from Chalk, Hagan-Burke, and Burke (2005).

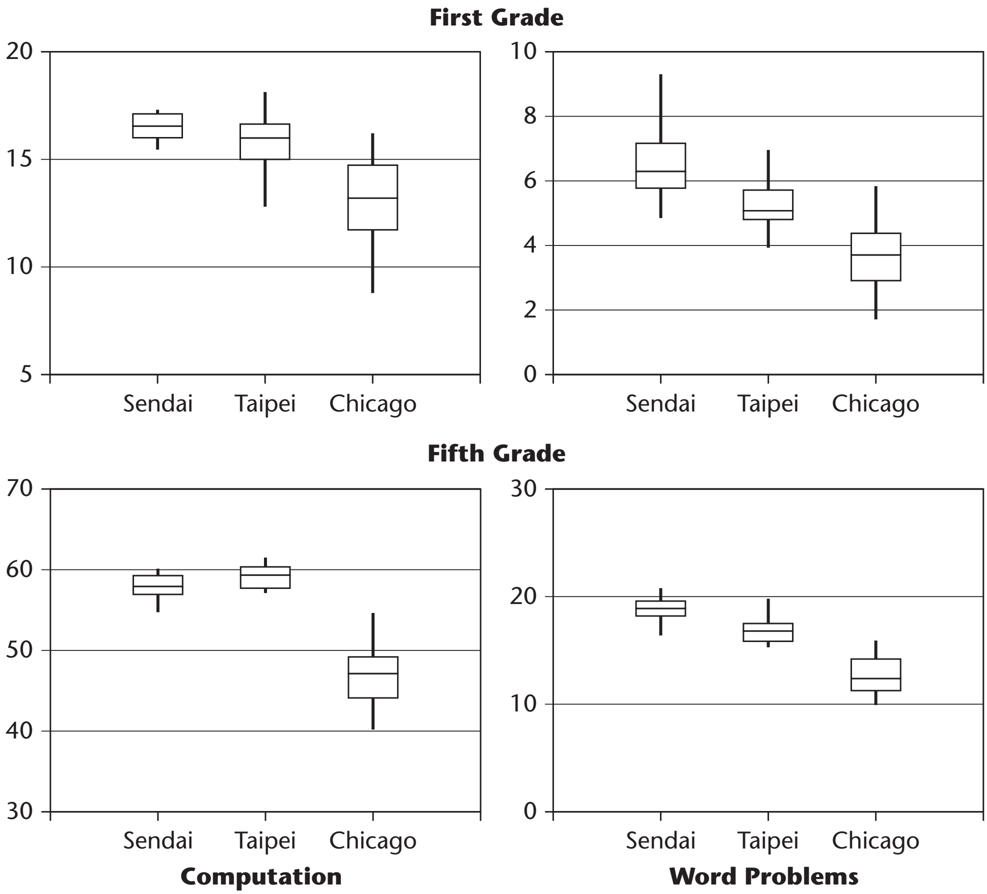
with learning disabilities. The researchers calculated the mean number of words written by students during three baseline periods and during three treatment periods in which self-regulated strategies were taught. They also calculated the standard deviations during each period. Table 3.5 shows how the mean number of words written increased from baseline to treatment conditions; the variability in number of words written (the standard deviations) remained somewhat similar throughout baseline and treatment conditions.

An unusual graph used to interpret measures of central tendency and variability is the box-and-whisker graph represented in Figure 3.7. This figure shows the mean number of correct answers on three mathematics tests in schools in Japan, Taiwan, and the United States for first and fifth grades. Each country's scores are displayed in a box with a vertical line at each end (the whiskers). The means are represented by the lines in the middle of the boxes. The whiskers show the full distribution of scores. The box shows the scores that fall between the 25th and 75th percentiles. The results in Figure 3.7 suggest that U.S. students have much greater variability in their scores than children in Japan and Taiwan and also have a lower average score in each area tested. Although there are many possible reasons for these differences, the figure conveys a lot of information quickly about the performance in each country and how it changes from first to fifth grade.

Measures of Relationship

There are many times when researchers and practitioners are interested in examining more than one variable at the same time. Are the number of hours spent doing homework related to student achievement? How does student self-esteem relate to achievement? Both of these examples involve the consideration of multiple variables in the same study; hours doing homework and achievement in question 1 and self-esteem and achievement in question 2. Let's say a researcher

FIGURE 3.7 "Box-and-Whisker" Graphs



Source: Adapted from Stigler and Perry (1990).

was interested in determining if there was a relationship between athletes' performances on the long jump and the high jump. We might enter this data in a table such as Table 3.6.

These data show the performance of Olympic gold medal winners in the high jump and the long jump. Note that each athlete has two scores, one for the long jump and one for the high jump. Looking at these data, can you see any relationship between the two scores?

It does seem that longer long jumps go along with higher high jumps, because both sets of numbers increase as we go from athlete 1 to athlete 22. However, a

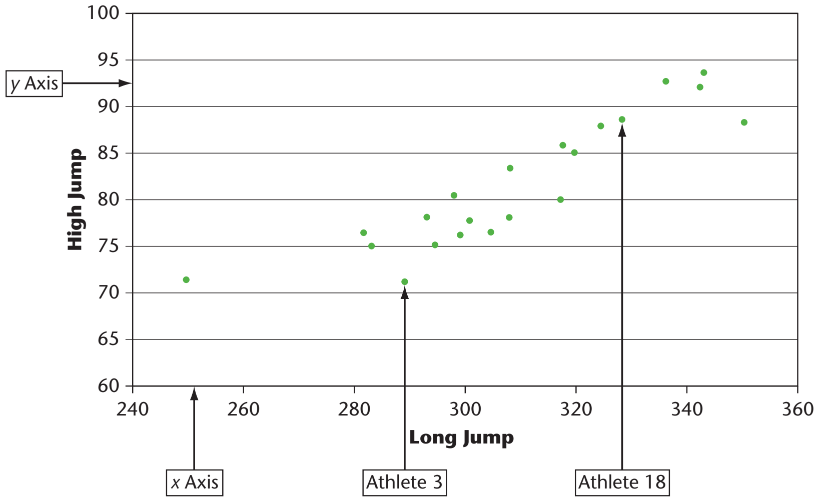
TABLE 3.6 Sample Table Showing Relationship Between Two Variables

Athlete Number	Long Jump (inches)	High Jump (inches)
1	249.75	71.25
2	282.87	74.80
3	289.00	71.00
4	294.50	75.00
5	299.25	76.00
6	281.50	76.25
7	293.12	78.00
8	304.75	76.38
9	300.75	77.63
10	317.31	79.94
11	308.00	78.00
12	298.00	80.32
13	308.25	83.25
14	319.75	85.00
15	317.75	85.75
16	350.50	88.25
17	324.50	87.75
18	328.50	88.50
19	336.25	92.75
20	336.25	92.50
21	343.25	93.50
22	342.50	92.00

Source: Adapted from Boggs, R. (n.d.). Retrieved from http://exploringdata.cqu.edu.au/datasets/oly_gold.xls

better way to see the relationship between two variables is through a graph known as a **scatterplot**. A scatterplot has an *x* and *y* axis with each axis representing one of the variables. In a scatterplot, each person in the study is represented by one data point on the graph. Values of one of the variables are plotted using the vertical or *y* axis of the graph, and values of the second variable are plotted using the horizontal or *x* axis of the graph. Each point represents the score for one person on both variables 1 and 2. Figure 3.8 displays the data from Table 3.6 in

FIGURE 3.8 Relationship Between High-Jump and Long-Jump Scores

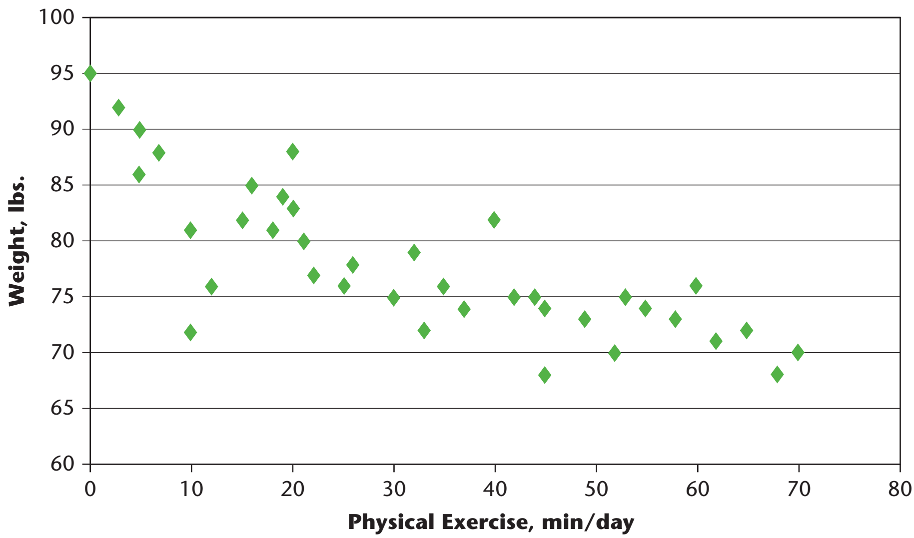


a scatterplot. In this case, the scores for the high jump are plotted on the *y* axis, and the scores for the long jump are plotted on the *x* axis. Note that each point in the graph represents one of the athletes. The data points representing athletes 3 and 18 are indicated.

With the use of the scatterplot, we can examine the patterns within the overall group to determine both the direction and the strength of the relationship or correlation. Remember that each data point represents two scores for each person, so as we move from left to right, we see how both variables are changing. If the data points tend to cluster in an upward pattern as we go from the left to the right side of the graph, this suggests that as high-jump scores go up, long-jump scores also go up. We call this a *positive correlation*. On the other hand, if the data points tend to go down as we move from the left to the right, this suggests a *negative correlation*. If we plotted data on physical exercise and weight for children age ten years, we might expect there to be a negative relationship or a negative correlation (as one variable goes up, the other goes down). This is shown in Figure 3.9.

So far, we have discussed how a scatterplot shows us the direction of the relationship between two variables. However, it also depicts the strength of the relationship. If the data points are scattered in an inconsistent way, the relationship is said to be *weak*. If the data points tend to be grouped in a narrower pattern

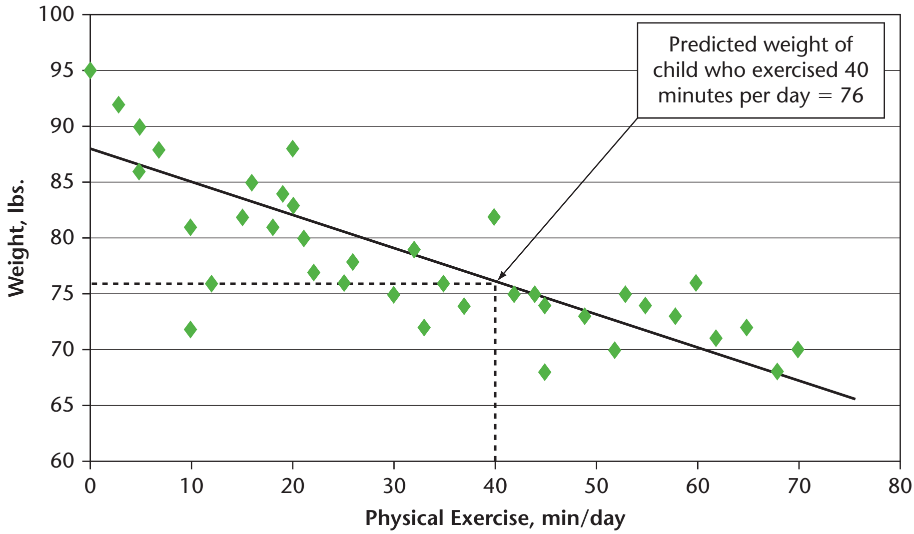
FIGURE 3.9 Scatterplot of Children's Weight and Amount of Physical Exercise



along a line, the relationship is said to be moderately *strong*. This is because the association is more systematic; as one variable changes, the other changes in a way that is visually predictable. In other words, if you were asked to predict the weight of any 10-year-old child who engaged in physical exercise 40 minutes a day, you could make a rough estimate based on the graph in Figure 3.9. Researchers make these predictions by plotting a *prediction line* that is calculated using a statistical formula that plots the line so that it is as close as possible to all of the points in the scatterplot. Predictions for individuals are made by locating the value of one variable (in this case, exercise) on the prediction line and reading the corresponding value of the other predicted variable (in this case, body weight). See Figure 3.10 for an example of a prediction line and how it would be used to predict the body weight of a child who exercised 40 minutes per day.

The direction and strength of a correlation can also be summarized numerically using a **correlation coefficient**. A correlation coefficient is a number that can be as large as +1.00 or as low as −1.00, but usually it is a decimal somewhere between these two numbers (for example, +0.65 or −0.48). The size of the number indicates how strong the correlation is, and the plus or minus sign indicates the direction (whether the relationship is positive or negative). As noted in the introductory discussion of correlational research in chapter 2, a large number shows a strong relationship regardless of whether the coefficient is positive or negative. A positive correlation is not better or stronger than a negative

FIGURE 3.10 Scatterplot with Prediction Line

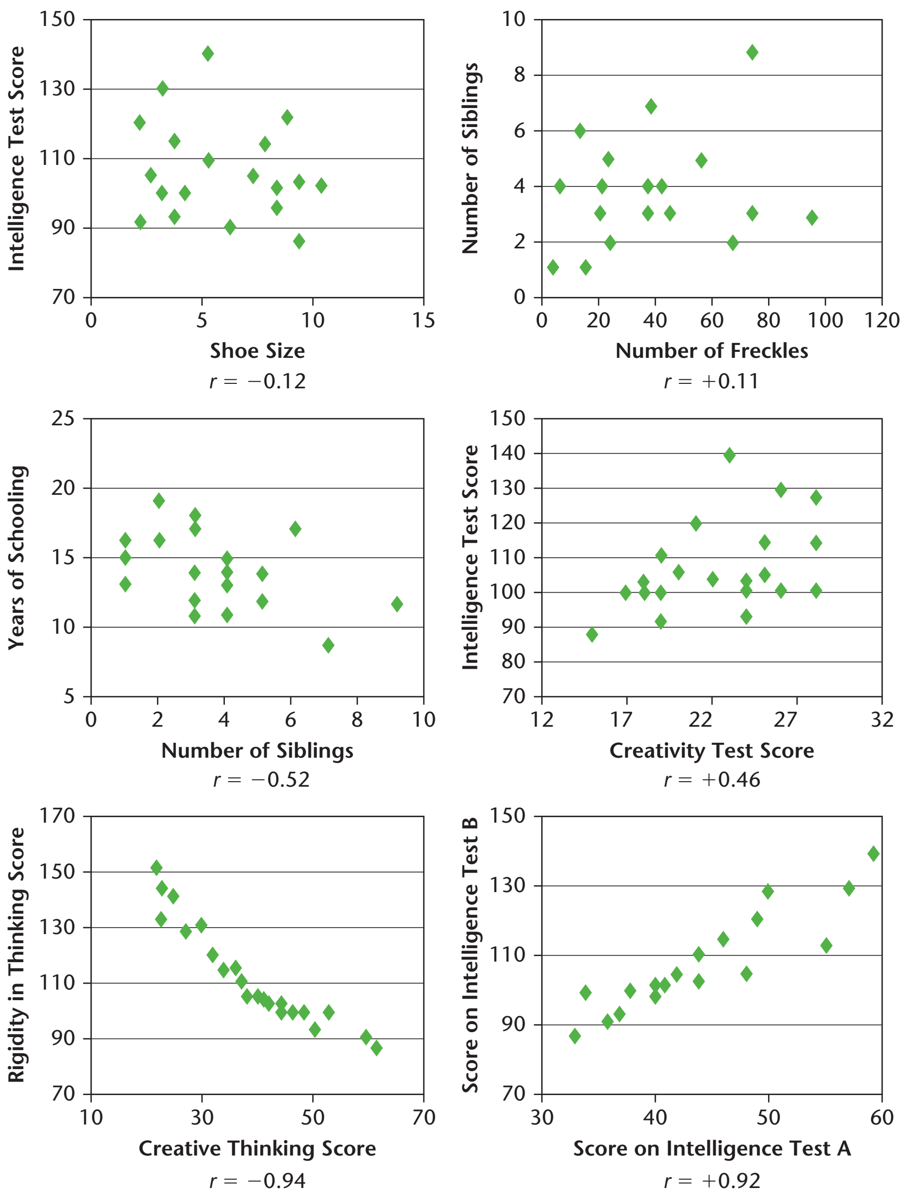


correlation; it simply indicates that the variables move in the same direction (as one goes up, the other also goes up) whereas, in a negative relationship, variables move in opposite directions (an increase in one variable is accompanied by a decrease in the other variable).

Although the direction, positive or negative, usually makes sense, for students who have not taken a statistics course, the concept of a correlation coefficient (the numerical portion) may be mysterious. Note that the coefficient does not represent the percentage of times the scores are related—a common misconception! The coefficient indicates the degree, or extent, to which the variables covary or go together. The best way to see this is in a scatterplot. Figure 3.11 shows the scatterplots for correlations of different strengths.

Note that the scatterplots in which data points are closer together along a line tend to have higher correlation coefficients, regardless of whether they are positive or negative. Scatterplots in which scores do not fall closely along a line show weak relationships and have lower numerical coefficients (again, regardless of whether they are positive or negative). A weak correlation means that there is some tendency for the variables to go together, but there is a lot of variability in the extent to which this occurs. If television watching and weight are weakly correlated, there will be a few children who watch little television but are overweight and others who watch a lot of television and are slender. This shows up in the scatterplot as widely scattered data points and in the correlation coefficient as a smaller numerical value. So, the size of the correlation coefficient gives us an

FIGURE 3.11 Scatterplot Representations of Correlation Coefficients of Different Sizes



overall estimate of the extent to which scores on two variables will be related in a way that is consistent across one large group of people.

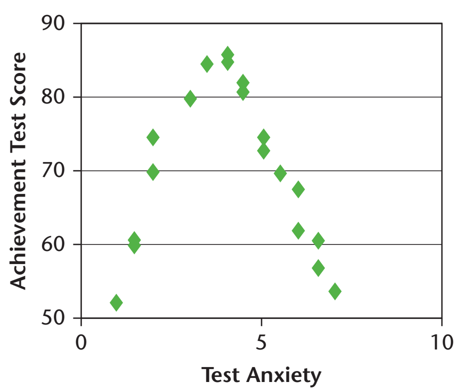
Correlational relationships can also differ in whether they are linear or curvilinear. A **linear relationship** means that as one variable changes, the other variable changes by an amount that is predictable using a straight line as the prediction line. In a **curvilinear relationship**, the direction of the relationship may be different at different levels of the variables. For example, the relationship between test anxiety and test scores is often curvilinear. A little anxiety is good because it gets one to study, but high levels of anxiety may interfere with learning or test taking, as depicted in the scatterplot in Figure 3.12.

To further your understanding of correlation coefficients, we demonstrate the way the correlation coefficients are calculated. One of the most commonly used correlation coefficients is the Pearson product-moment correlation coefficient (**Pearson *r***). The Pearson r is calculated when the two variables are at least interval data. The calculation formula for Pearson r looks complicated, but calculation of each component of the formula is relatively easy.

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{\left(\sum X\right)^2}{N}\right)\left(\sum Y^2 - \frac{\left(\sum Y\right)^2}{N}\right)}}$$

Let's use the example of self-esteem and achievement discussed earlier in the chapter. For the sake of this example, self-esteem will be the X variable and

FIGURE 3.12 Scatterplot of Relationship Between Achievement Test Scores and Test Anxiety



achievement the *y* variable. The researcher gives each student a measure of self-esteem and a measure of achievement and generates the following paired scores:

Student	X	Y
1	120	90
2	100	75
3	109	80
4	65	54
5	75	95
6	107	92

First you will see that we need the mean and standard deviation for the *X* distribution (self-esteem scores) and the *y* distribution (achievement). The mean of the *X* distribution is equal to 96 and the mean of the *y* distribution is equal to 81. Using the standard deviation formula presented earlier, the standard deviation of the *X* distribution is equal to 21.37 and the standard deviation of the *y* distribution is equal to 15.26. (You might want to practice your standard deviation calculation with this example!)

Let's set up a table with all the needed components. Be prepared for some very large numbers!

First, sum (Σ) all of the *X* scores (you will have this number from your calculation of the mean). The sum of the *X* scores is equal to 576.

Second, square each *X* score and then sum all of the X^2 scores. The sum of the X^2 scores in our example is equal to 57,580.

Third, sum the *Y* scores. The sum of *Y* is equal to 486.

Fourth, square each *Y* score and then sum all of the Y^2 scores. The sum of Y^2 in our example is equal to 40,530.

Fifth, calculate the product of *X* times *Y* or the cross products, and sum all of the cross products. The sum of the cross products is equal to 47,499.

Student	X	X^2	Y	Y^2	(XY)
1	120	14400	90	8100	10800
2	100	10000	75	5625	7500
3	109	11881	80	6400	8720
4	65	4225	54	2916	3510
5	75	5625	95	9025	7125
6	107	11449	92	8464	9844
$n = 6$	$\Sigma X = 576$	$\Sigma X^2 = 57580$	$\Sigma Y = 486$	$\Sigma Y^2 = 40530$	$\Sigma XY = 47499$

Now you have all of the essential components for the Pearson r formula. The calculation is shown in Exhibit 3.1. For our example, $r = .517$.

EXHIBIT 3.1
Sample Pearson r Correlation

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{\left(\sum X\right)^2}{N}\right)\left(\sum Y^2 - \frac{\left(\sum Y\right)^2}{N}\right)}}$$

$$r = \frac{47499 - \frac{(576)(486)}{6}}{\sqrt{\left(57580 - \frac{(576)^2}{6}\right)\left(40530 - \frac{(486)^2}{6}\right)}}$$

$$r = \frac{47499 - 46656}{\sqrt{(57580 - 55296)(40530 - 39366)}}$$

$$r = \frac{843}{\sqrt{(2284)(1164)}}$$

$$r = \frac{843}{\sqrt{2658576}}$$

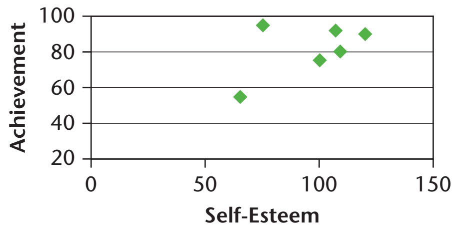
$$r = \frac{843}{1630.5}$$

$$r = .517$$

The moderate and positive correlation is plotted in Figure 3.13. A correlation of +.517 indicates that self-esteem and achievement are positively related. As self-esteem improves, so does achievement. The small size of the correlation in this example is likely a function of the small sample size.

Studying self-esteem and achievement is an example of a bivariate correlation. A bivariate study is one in which only two variables are included in a study. The long-standing and well-accepted tenet of correlational research is that *correlation does not imply causality*. This is certainly true when one is examining a

FIGURE 3.13 Relationship Between Achievement and Self-Esteem



bivariate correlational study. Consider achievement as an example. While our study showed a positive relationship between self-esteem and achievement, this does not mean that change in self-esteem causes change in achievement or vice versa. Think about all of the other variables that could influence achievement. We could never conclude based upon our bivariate study that improved self-esteem causes improved achievement.

Some more advanced correlational procedures, such as multiple regression, include multiple variables. Under these conditions some researchers may conclude causality. Note that when reading studies that utilize multiple regression analysis, always examine the controls utilized by the researchers. Statistical and procedural control of extraneous variables is an essential to causal interpretation. These concepts are discussed in greater detail in chapter 10.

SUMMARY

Numerical data gathered by researchers and practitioners are summarized and made meaningful by using descriptive statistics. Prior to deciding how to analyze quantitative data, researchers must determine the scale of measurement to be used. These scales form a hierarchy from the simplest to the most sophisticated data, for example, from nominal to ratio. Determining the scale of measurement will dictate the type of statistics that can be used to analyze the data.

There are several basic statistics used to analyze and interpret quantitative data. Data can be summarized by constructing a frequency distribution and then graphing the data using either a frequency polygon or a histogram. Additionally, most educational professionals summarize their data using measures of central tendency and measures of variability. The mean is believed to be the best measure of central tendency for interval or ratio data, since it is sensitive to every score in the distribution. The median and mode are also used as measures of central tendency.

Variability in a distribution is demonstrated through the calculation of the variance and standard deviation of the distribution. The standard deviation provides important information about how the scores in the distribution vary around the mean and the spread of the scores.

Data can form a variety of distributions. The most commonly known distribution is a normal distribution. In a normal distribution, the mean, median, and mode are the same number. The distribution is symmetrical. If we folded the distribution in half, the result would be two equal halves. The number of scores decreases as the scores move away from the mean. The normal curve has certain useful mathematical properties. One property that is useful is that a certain percentage of scores fall between the mean and certain distances above and below the mean. These distances are described as how many standard deviations above or below the mean a score falls. Approximately 68% of the scores fall between plus and minus one standard deviation from the mean. Approximately 95% of the scores fall between plus and minus two standard deviations from the mean. Ninety-nine percent of the scores fall between plus and minus three standard deviations from the mean.

Distributions, however, are not always normal. Skewed distributions are not symmetrical. In a negatively skewed distribution, most of the scores are on the high end, and in a positively skewed distribution, most of the scores are on the low end of the distribution.

When educational professionals examine relationships between two or more variables, a correlation coefficient is calculated. The Pearson product-moment correlation coefficient (Pearson *r*) is calculated when the variables are on an interval scale. The correlation coefficient (*r* value) can only assume values between negative one and positive one. The higher the absolute value of the correlation coefficient, the stronger the degree of relationship. Regardless of the strength of the relationship, conclusions about causality are not acceptable. The correlation coefficient can be positive or negative. If the *r* value is positive the increase in one variable is accompanied by an increase in the other variable or a decrease in one variable is accompanied by a decrease in the other variable. A negative *r* value indicates an inverse relationship, for example, as one variable increases the other variable decreases. Correlations are often graphed using scatterplots.

In conclusion, descriptive statistics serve an important function in research and in educational practices. They provide a basis for understanding, interpreting, and representing data.

KEY CONCEPTS

bimodal distributions
correlation coefficient
curvilinear relationship
descriptive statistics
deviation score
frequency distribution
frequency polygon
frequency table
histogram
linear relationship
mean
measures of central tendency

median
mode
negatively skewed distribution
normal curve
nominal, ordinal, interval, and ratio scales
normal distribution
outlier scores
Pearson *r*
positively skewed distribution
quantitative data
range
scatterplot
standard deviation
sum of squares
variance

DISCUSSION QUESTIONS OR ACTIVITIES

1. Consider the data in Table 3.7.
    a. Calculate the mean of each distribution.
    b. Determine the median and the mode of each distribution.
    c. Calculate the variance and standard deviation of each distribution.
    d. In which distribution is there more variability? How do you know?
2. You are interested in knowing if the students' scores on the math tests are related to their scores on the ELA test.
    a. Create a scatterplot; make the math scores the *x* axis and the ELA scores the *y* axis.
    b. What does this tell you about the relationship between the scores?
    c. Calculate the Pearson *r*.
    d. How does the actual correlation compare with your answer to (b) above?

TABLE 3.7 Student Math and ELA Scores

Student Number	Math Scores	ELA (English Language Arts Scores)
Student 1	80	75
Student 2	75	85
Student 3	73	90
Student 4	96	95
Student 5	80	65
Student 6	45	60
Student 7	66	60
Student 8	96	99
Student 9	73	99
Student 10	77	50

TABLE 3.8 Correlations Among Five Variables

	Homework Assignments Completed	Total Teacher Science Grades for Three Semesters	Total Teacher Ratings on Science-Related Skills	State Science Test Raw Score Criterion for Passing Score = 65
Days Absent	−.25	−.33	−.56	−.60
Homework Assignments Completed		+.48	+.38	+.42
Total Teacher Science Grades for 3 Semesters			+.45	+.79
Total Teacher Ratings of Science-Related Skills				+.53

3. Consider Table 3.8 and examine the correlation coefficients presented.
   a. Which variable is the strongest predictor of state science test scores?
   b. Which variable is most strongly related to Teacher Ratings of Science-Related Skills?
   c. Which pairs of variables show the weakest relationships?
   d. Can we conclude that improving attendance will increase state Science Test scores? Why or why not?

SUGGESTED READINGS

Holcomb, E. L. (1999). *Getting excited about data: How to combine people, passion, and proof.* Thousand Oaks, CA: Corwin.

CHAPTER FOUR

EDUCATIONAL MEASUREMENT

Archival Data and Standardized Instruments

CHAPTER OBJECTIVES

- Define archival data and give three examples of archival data found in a typical school setting
- Explain some of the reliability issues associated with archival data
- Describe the characteristics of standardized tests
- Distinguish among norm-referenced, criterion-referenced, and self-referenced instruments
- Describe the cognitive and noncognitive domains measured by standardized tests and be able to provide an example of each
- Interpret the different ways standardized scores are reported
- Explain the difference between reliability and validity and give several examples for each
- Know the criteria for selecting a standardized test
- Understand the principle of assessment accommodations
- Be familiar with issues related to bias in testing

MEASUREMENT IN EDUCATION

In an era of accountability, measuring educational outcomes has come under careful scrutiny. Federal and local governments, school boards, communities, and parents demand evidence that students are learning and meeting agreed-upon standards. Educational professionals examine research and *best practices* to determine how to measure student learning outcomes and use assessment to improve educational instruction and decision making. Both research and best practice often recommend the utilization of archival data and standardized instruments as one way to measure and understand student learning, student behavior, and the learning context. These concepts are the focus of this chapter.

However, there are many times when practitioners and researchers may decide to develop their own instruments and not use standardized instruments. The following are examples of some of the reasons why this might occur:

- It may be that a standardized instrument is not appropriate for the variables being studied. It may be that the practitioner or researcher wants an instrument directly related to the context and the participants being studied.
- There may not be a standardized instrument that measures the variable under consideration. One student recently wanted to collect quantitative data on cyberbullying. She hoped to find a standardized instrument to use for her master's thesis. After searching the literature, she could not find such an instrument. Her only option was to use the literature to develop her own instrument to measure cyberbullying.
- A standardized instrument exists, but it does not measure the variable in the exact way the researcher or practitioner designs the study or assessment. In this case, the researcher may modify the standardized instrument to better meet his or her individual needs.
- There is no standardized instrument that was developed using a sample like the sample chosen for the study. In this case, if the researcher used the instrument, the interpretation of the results would not be appropriate.

In these cases, the researcher and practitioner would develop their own quantitative instruments. These instruments could measure a variety of variables. This chapter is devoted to archival and standardized instruments. Discussion of nonstandardized instruments can be found in chapters 8 and 12.

Archival Data

Archival data are data that have already been collected, typically by an individual teacher, school, or district rather than by a researcher. There are many types of archival data, but in educational research this typically includes data a school might keep at the individual student level, at grade level, and at building and district levels. Student absenteeism, graduation rates, suspensions, standardized state test scores, and teacher grade-book data are examples of archival data that might be used in educational research or assessment. It is not uncommon for researchers to combine several different types of archival data in a single study.

Parents, educators, researchers, and the general public have greater access to archival data on schools and school districts because of some of the regulations embedded in the No Child Left Behind (NCLB) legislation. This archival data is reported in the form of *school report cards*. School report cards are often available on state departments of education Web sites. While state report cards vary in the way the data are reported, most include the same information. What follows are brief examples of the type of information that is often included in these reports. It should be noted here that one of the requirements under NCLB is that all information must be **disaggregated**, which means that the data is reported by subgroups. The subgroups typically reported include race, ethnicity, gender, students with disabilities, English language learners, and students' socioeconomic status. Further, report cards often include multiple years so that trends may be analyzed and comparison to other schools and districts in the state can be made.

- *Enrollment and average class size.* This category provides total enrollment number as well as enrollment by class.
- *Demographics.* This category minimally includes the total number of students by subgroups, the number of students receiving free or reduced lunch, the number of students for whom English is a second language, and so on.
- *Attendance and suspension rates.* Total attendance and suspension rates are provided by grade and subgroups.
- *Teacher qualifications.* This category provides detailed information on the degrees held by teachers as well as the number of teachers teaching outside of their certification areas.
- *Graduation rates.* This may be one of the most controversial categories; the calculation of graduation rates may vary by state and district.
- *Achievement scores.* This category includes the actual scores on state and national tests as well as whether a school is meeting its AYP (Adequate Yearly Progress under NCLB).
- *Postsecondary plans for students.* This includes the number of students who upon graduation plan to go directly into the workforce or attend vocational institutions or two-year and four-year colleges.
- *Budget information.* The information provided is typically per pupil expenditure and total operating budget.

It is important to review your own state's way of reporting this information because, as indicated, it varies by state. Researchers, practitioners, and the general public will find that becoming familiar with school and district report cards is helpful. For researchers, school report cards help to put their research into a context. If the research is conducted in a school, is the school meeting its AYP? Are most of the students receiving free or reduced lunch? For the practitioner, the report card provides a way to compare student performance to the performance of similar students in other schools. For the general public, report cards may influence where families purchase homes so that children might go to districts with high levels of funding and high levels of academic achievement.

While archival data can reveal a great deal about schools, school districts and their students, there are concerns about the accuracy of these data and the extent to which high-stakes testing may have led some schools to distort their data.

BOX 4.1
What Is the Real Graduation Rate for U.S. Students?

With the emphasis on educational accountability, graduation rates have come under increasing scrutiny. Under NCLB, state accountability systems are required to include high school graduation rates as one of their accountability indicators. It might seem that graduation rates would be relatively simple to calculate: you divide the number of students graduating by the total number of students. However, questions quickly arise about:

- How do you determine the total number of students? Is it the total number in the freshman class of a high school? Is it the total number in the junior class? How can you account for students who transfer or move? When and where should students be counted? How do you determine when and if a student has dropped out?
- Does it matter when and how students graduate? If a student takes five or six years to graduate, should that student be included in the graduation totals? If students earn a General Equivalent Development certificate, are they included in the graduation totals? If students leave school without graduating and then return to complete their education, should they be counted?

Swanson (2003) examined many of these issues by calculating graduation rates for states and the nation as a whole using three different methods. One method, developed by the National Center for Education Statistics (NCES), compared the number of students graduating in a given year with the number of students enrolled for that year plus the number who dropped out in the three previous years. Swanson pointed

(See Box 4.1 for a discussion of this problem in relation to graduation rates.) In addition, the methods used to summarize and present archival data can often result in a distorted picture of a school's performance, a problem that is illustrated in the next section of the chapter.

Reliability and Validity in Archival Data One of the benefits of using archival data is the amount of time it saves. Having the data already collected speeds up a research study considerably, and for this reason, archival data are a favorite of graduate students. However, researchers should proceed with caution when using archival data. Despite their attractiveness, one barrier to archival data is that they have been collected by someone else, and therefore, the researcher cannot have any quality control over their collection. Procedures for collecting data might

out that dropout rates are often underreported and data are not available for some states. The overall graduation rate calculated using the NCES method was 85%, which is the rate frequently reported as the national rate. However, the two other methods of calculating graduation rates that did not use dropout rates produced estimates that were at least 10% lower. Both of these estimates used student enrollment or promotion rate in previous years as the basis for their estimates. State graduation rates also differed using the different methods, with some states having graduation rates as low as 48.6%. In an effort to further determine the best estimates of graduation rates, Swanson (2003) reports that research from the Urban Institute suggests that using data reported from public school districts, graduation rates in the United States may be alarmingly as low as 68%. When these findings are carefully analyzed by subgroups, Swanson reports that only a little more than 50% of students from disadvantaged racial and ethnic groups leave high school with a diploma.

To further complicate the matter of how to calculate graduation rates, some districts have been accused of "using deceptive practices to mask the true extent of their dropout problems" (Swanson, 2003, p. 4). According to Swanson, more than a dozen schools in the Houston Independent School District were accused of underreporting their dropout rates and falsifying their records. Furthermore, the New York City public schools were accused of encouraging poor-performing students to enroll in alternative educational programs. These students, many of whom would not likely receive a high school diploma, remained on the school district enrollment numbers, thereby boosting the district's graduation rates.

NCLB has provided some guidelines for reporting graduation rates, yet the way districts calculate and report these rates still vary greatly. According to NCLB guidelines, students who complete high school within a "standard number of years" (Swanson, 2003, p. 12) should be part of the graduation rates calculation.

differ from person to person, resulting in poor reliability, or data collection may be inaccurate, raising questions about validity.

For example, say that a researcher is investigating the impact of a building-wide behavior program. School building A adopted a behavior modification program the previous year, whereas school building B (similar to A in all its demographics) did not implement the program. The archival data for this study are both buildings' suspension records. To establish baseline data, the researcher took three years of suspension records from both buildings and determined that both buildings averaged consistently about 300 suspensions each year. Then the researcher used records to examine the number of suspensions during the previous academic year, when school building A implemented its program. In this review of the records, the researcher found that school building A decreased the number of suspensions over the previous year by 50%, whereas school building B continued to have about 300 suspensions. Sounds like the behavior modification program worked, right? Not necessarily. Inaccuracies in the data from school building A might be distorting the results. For example, did school building A, as part of the new program it implemented, revise its criteria for suspension? Let us say that it did, and that acts committed by students that were used to get them suspended now are handled by the teacher in the classroom, thus lowering the number of incidents reported. In other words, the new criteria simply allowed school building A to report fewer suspensions, because of the new criteria, and undermined the reliability of the data. It did not mean that the new program worked to prevent suspensions. However, because the researcher was not a part of the data collection process and only used such data later on to perform the study, the researcher probably would not know that such a change in policy occurred. Other types of archival data, such as percentage of student body who receive free or reduced-fee lunch, high school dropout rates, and even graduation rates, are susceptible to inaccuracies.

When working with archival data, it is important to consider possible inaccuracies that the data may contain. If possible, a researcher should validate or double-check how the raw data have been collected. To do so, the researcher might conduct interviews with the individuals who originally collected the data. The purpose of these interviews would be to gather a better understanding of the methods used in collecting the data and to investigate any changes in policy that might influence the archival data. If possible, the researcher would also want to examine the instruments used to collect the data and any information about their development and administration. Often such scrutiny of the original data or instruments is not possible. If the archival data have so many flaws that they jeopardize the validity of the study's results, then archival data should be avoided altogether, and another method for collecting the data should be employed.

Standardized Instruments

Inaccuracies in archival data result in many researchers turning to standardized instruments to supplement their data collection. **Standardized instruments** refer to a category of measuring tools that have already been developed and piloted, usually by someone other than the researcher who is doing the current study. These instruments may include tests, observational rating scales, questionnaires, or even scoring protocols for interviews. A standardized instrument is defined as an instrument with the following characteristics:

- It includes a fixed set of questions or stimuli.
- It is given in a fixed time frame under similar conditions with a fixed set of instructions and identified responses.
- It is created to measure specific outcomes and is subjected to extensive research development and review.
- Performance on the instrument can be compared to a referent such as a norm group, a standard or criterion, or an individual's own performance as described following.

Types of Standardized Instruments Standardized instruments are typically designed by researchers at universities or teams with expertise (called **psychometricians**) in the area of instrument development who work for test-publishing companies. Once developed, the tests and other measures are used by researchers and practitioners in both conducting research and measuring school outcomes. There are also researchers who develop their own standardized instruments to measure the variables in their research studies. Standardized tests are typically administered under standard procedures and are scored according to standard instructions.

Standardized instruments are divided into three main categories. **Norm-referenced instruments** are tests in which the scoring indicates how a student's performance compares with that of a **norm group** (that is, a group of participants who have already taken the instrument and their scores represent the range of possible performance on the measure for an identified population). The Iowa Test of Basic Skills (ITBS) is an example of a norm-referenced test that is frequently used in educational research and assessment. Test results for individual children or schools may be compared with those of a variety of norm groups. The test includes separate norms for students from subgroups such as:

- High- and low-socioeconomic areas
- Private religious and public schools

- Urban areas and nonurban areas
- English speakers and non-English speakers (Brookhart, n.d.)

Criterion-referenced instruments are tests in which the scoring involves a comparison against a predetermined standard of performance or a criterion. For example, many states define the range of scores on the state assessment tests that qualify as inadequate, passing, or excellent performance. These tests are often developed using learning standards that serve as the criterion for performance and are often referred to as *standards-based tests*. **Self-referenced instruments** are tests that measure an individual student's performance over time to see if it improves or declines when compared with past performance. Some standardized instruments provide information that allows both norm-referencing and self-referencing. For example, the Dynamic Indicators of Basic Early Literary Skills is a test designed to allow measurement of individual student progress over time *and* comparisons to the norm group (University of Oregon, 2004).

BOX 4.2
High-Stakes State-Mandated Tests

High-stakes tests are tests that are used to make "high-stakes" decisions. These decisions often include whether a student is promoted to the next grade or permitted to graduate. State-mandated tests are certainly not new. For years, many states used mandated tests to assess student performance in the content areas. However, with the passage of the No Child Left Behind Act, states are now mandated to create their own standards for student achievement in order to demonstrate that students are making adequate progress to meet those standards.

The standards developed by the states form the basis for the tests and have become part of curricular and instructional decision making. Teachers are strongly encouraged to incorporate the standards into their instruction. From a test construction point of view, most of the tests are multiple choice or essay, and a few states have performance or portfolio assessments. Criterion-referenced scoring is used in most of the states where student performance is compared with some standardized benchmark that indicates level of performance or scores that are considered passing.

The use of high-stakes state-mandated tests is not without controversy. Many policymakers and government officials believe that these tests will result in positive outcomes related to student performance and quality of education. Generally, advocates of high-stakes state-mandated testing believe that student academic performance will increase; more time will be spent teaching content; and poor-performing schools will see improvements in teaching and learning. Because high-stakes tests report the results for all students, students who have been traditionally

Although there is a wide range in what standardized instruments measure, the processes associated with their development are similar across instruments. It is important to understand that much time, energy, and expertise go into establishing a standardized instrument. It is common for these measures to undergo years of work in their creation, plus continuous refinement to ensure their accuracy and reliability. There is considerable controversy surrounding the use of standardized tests in education, especially those involved in high-stakes testing. Box 4.2 summarizes some of the issues involved in this debate.

Cognitive and Noncognitive Areas Measured by Standardized Instruments

In general, standardized instruments measure five broad areas: achievement, aptitude, personality, attitudes or interest, and behaviors. Note any of these categories can include measures that are norm-referenced, criterion-referenced, or self-referenced. Achievement and aptitude are examples of **cognitive measures**.

underserved (ethnic minorities, the poor, and students with disabilities) will now have their performance monitored more closely (Linn, 2003).

Critics argue that the consequences of these tests will be negative. Among their concerns are that teachers feeling pressured by administrators and the community to improve student performance will "teach to the test." They will spend more time on test preparation than teaching content. Furthermore, the time that was devoted to further exploration of content in a creative way will be eliminated because of time considerations. Another concern voiced by critics is that the state tests are focused on a narrow set of skills and knowledge rather than higher-level cognitive skills. If this is in fact the case, the curriculum presented to students will be narrowed and focused on lower-level cognitive skills such as memorization and rote knowledge. An equally important concern is that many of the high-stakes state-mandated tests discriminate against low-socioeconomic status and ethnic minority children. A frequently asked question is whether test questions measure only content taught in schools or if they measure knowledge acquired outside of school. If tests do include knowledge gained by experiences outside of school, then ethnic minorities and children from low-income families are at a disadvantage. Moreover, some research (Darling-Hammond, 2001) has suggested that because of the intense pressure to improve performance, some schools have resorted to pushing low-scoring students into special education, retaining them in their current grade levels, or even encouraging them to drop out of school!

The research on this topic is limited and inconsistent yet vital. However, given the climate of accountability and the regulations of NCLB, further research is essential.

Personality, attitude or interest, and behavior tests are examples of noncognitive measures. **Achievement tests** are generally associated with measuring what a student has already learned in school. They tend to measure content knowledge such as facts, concepts, or principles and skills such as computation, problem solving, listening, and writing. Most of the tests used in contemporary high-stakes testing are achievement tests. Many times these tests are developed as a test battery. A **test battery** is a collection of several subtests or mini-tests, each measuring a specific skill such as reading, mathematical literacy, or science. For example, the Iowa Test of Basic Skills includes subtests for vocabulary, comprehension, math computation, and problem solving, just to name a few.

Unlike achievement tests, **aptitude tests** are designed not to measure what someone knows, but to predict what one can do or how one will perform in the future. Intelligence tests and tests of mental or cognitive abilities are examples of aptitude tests. On an aptitude test, students are asked to demonstrate their ability to solve problems and apply knowledge and skills to solve problems and complete tasks. Ideally, successful performances should not depend on past experience. It should only be affected by the students' problem-solving abilities. Therefore, the questions on these tests present situations and tasks that are novel and one hopes unfamiliar to most students. If the questions are familiar, they should be familiar to all students and not a subgroup of students. Tests such as the Wechsler Scales and the Terra Nova are examples of commonly used aptitude tests.

Personality tests, an example of a noncognitive measure, differ from achievement and aptitude tests in that they measure self-perception or personal characteristics. Personality tests usually present the person with a set of questions intended to measure that individual's traits, characteristics, or behaviors. Personality is measured through objective tests such as the Minnesota Multiphasic Personality Inventory or the Myers-Briggs Type Indicator. These inventories require individuals to respond to a set of statements that reflect their personality traits. These types of tests are relatively simple to score in comparison with projective personality tests.

Projective personality tests require individuals to give verbal responses to what they see in unstructured and ambiguous stimuli. The Rorschach Inkblot Technique and the Thematic Apperception Test are examples of projective personality tests. Because these tests involve complex stimuli and responses, the person administering and scoring the tests must be well trained and highly skilled.

Another type of noncognitive measure commonly used in both research and practice is an **attitude** or **interest scale**. Like personality tests, these measures typically use self-report questions to assess a person's attitudes toward a topic or interests in areas. For example, career counselors use a variety of interest inventories to help students think about areas of work or study that they might pursue. One such inventory is the Campbell Interest and Skill Survey. This measure allows students to compare their self-reported interests and skills with those of professionals engaged in a wide variety of careers to see if they would be similar to people already in the field. Measures of attitudes toward a wide variety of educational issues have also been created by researchers. An example is the Mathematics Self-Efficacy Scale, a 34-item scale designed to measure one's "beliefs that he or she is capable of performing math-related tasks and behaviors" (Ciechalski, n.d., ¶2).

Curriculum-based measures (CBM) are a way of monitoring student progress through assessment of academic skills using curricular materials commonly used by classroom teachers. They are based upon the assumption that all children need to learn basic skills and that speed and fluency are good indicators of proficiency. CBMs are brief, timed questions or probes lasting approximately one to five minutes and are given with relative frequency. The measure is made up of materials taken from the student's school curriculum and are used as a way to monitor the student progress. Typical areas assessed include reading, mathematics, spelling, and written expression. CBMs allow education professionals to determine whether an academic intervention is improving student learning outcomes. CBMs are a widely practical measure that practitioners often find useful. If you are interested in more information, you can review the material connected with the DIBELS (Dynamic Indicators of Basic Early Literacy Skills), which can be found at http://www.dibels.uoregon.edu. Specific instructions for ways to use CBMs and calculate the effectiveness of interventions can be found at http://www.interventioncentral.org.

A final category of measures that is useful for both researchers and educators includes **behavior rating scales** and **checklists**. The rating scales are used to quantify observations of behaviors, often to assist in making diagnoses of problems. The rating scales might be completed by a parent, teacher, child, or other school staff. Checklists include lists of behaviors that are simply checked to indicate the occurrence of behaviors. Rating scales present numerical rating scales to assess the frequency or intensity of behaviors. For example, the Achenbach Behavior Rating Scale is often used by school psychologists to assess problem behaviors in students.

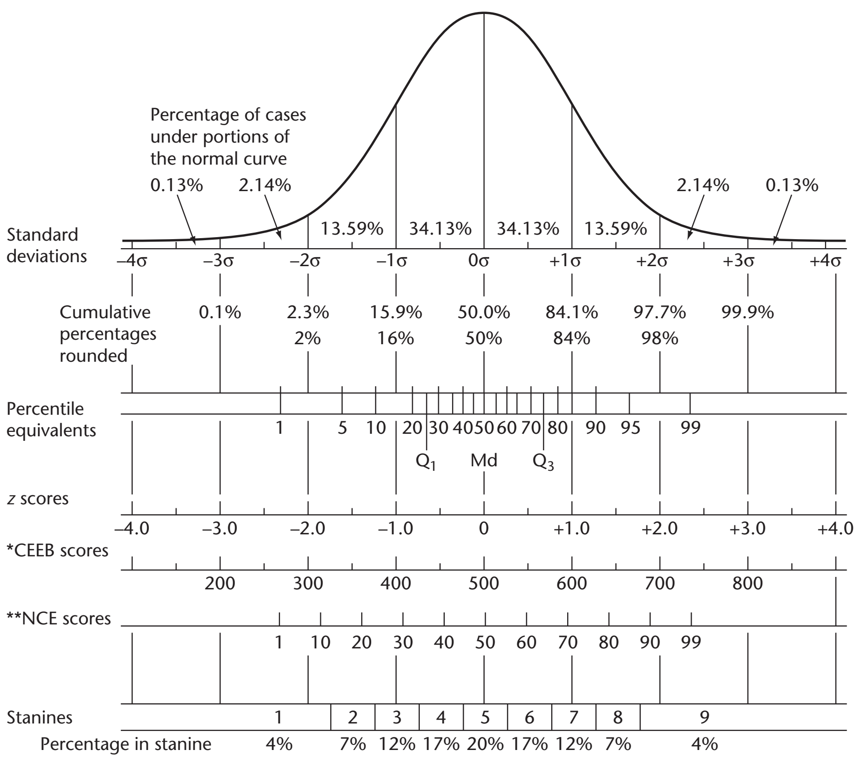
Types of Scores Used to Compare Performance on Standardized Tests

Several types of scores are used to compare performances on educational measures.

Raw Scores **Raw scores** are scores that summarize a person's performance on a measure by showing the number of responses or summing the scores for each response. Performance on classroom tests is usually reported as raw scores, but this type of score is not used by most standardized measures. The reason is that one cannot know what a raw score represents without knowing more about the mean and standard deviation (SD) of the distribution of scores.

Percentile Ranks **Percentile ranks** are scores that indicate the percentage of persons scoring at or below a given score. A percentile rank of 82 means that 82% of persons scored below that score. Figure 4.1 shows the percentile ranks that correspond to scores based on the number of standard deviations between the score and the mean in a normal distribution (note in Figure 4.1 the symbol used for standard deviation is a lower case sigma: σ). For example, reading down from the curve, one can see that a score that is one standard deviation above the mean has a percentile rank of approximately 84. A score that is .50 of a standard deviation below the mean has a percentile rank of approximately 31. You can calculate the exact percentile rank of a score if you know how many standard deviations it is above or below the mean. For a computer simulation that calculates the exact

FIGURE 4.1 Normal Curve with Test Scores



Note: *CEEB refers to College Entrance Exam Board test, for example, SAT; **NCE refers to Normal Curve Equivalent.

Source: Adapted from Seashore (1980).

percentile rank based on the normal curve, go to the Web site http://www.stat.berkeley.edu/users/stark/Java/NormHiLite.htm.

Standard Scores **Standard scores** are raw scores that have been normalized or converted to a standard scale. They indicate how far a score is from the mean of distribution in terms of standard deviation units. Many standardized measures report student scores as standard scores, allowing a direct comparison of student performance. A standard score shows immediately how well a person did in comparison with the rest of the group taking that measure. Examples of standard scores are *z* scores and college entrance exams (referred to as CEEB scores in Figure 4.1) such as the Scholastic Assessment Test (SAT) or Graduate Record Examinations (GRE).

The most basic standard score is the ***z* score**, which is calculated in the following manner:

$$z \text{ score} = (\text{score} - \text{mean})/\text{standard deviation}$$

Any score can be converted to a *z* score if the mean of a distribution and the standard deviation are known. If Karyn got a score of 75 on a test with a mean of 65 and a standard deviation of 5, her *z* score would be as follows:

$$z \text{ score} = (75 - 65)/5 = 10/5 = +2.0$$

As you can see, the *z* score represents simply the number of standard deviations away from the mean that the score is. Since all standard scores are based on the normal curve, knowing a person's *z* score would allow you to determine, for example, their percentile equivalents. Therefore, you can go to Figure 4.1 and determine Karyn's percentile rank, knowing that she is two standard deviations from the mean. Try one more example. Peter got a score of 55 on the same test. What would his *z* score be? What would his percentile rank be?

One difficulty in using *z* scores is the possibility of a negative number. How would you explain a *z* score of -1.00 to a parent? Because many people do not like using negative numbers, some measures convert their raw scores into slightly different standard scores. For example, SAT and GRE scores are simply *z* scores multiplied by 100 + 500. So they have a mean of 500 and a standard deviation of 100. Knowing this, you should be able to use Figure 4.1 to determine the percentage of people who score between 300 and 500 on the SAT. If you need a hint, this would be the area between one and two standard deviations below the mean.

Stanines **Stanine scores** are a type of standard score that divides a distribution into nine parts, each of which includes about one half of a standard deviation. Students are assigned scores ranging from 1 through 9 based on their performance. All students within a given stanine are considered to be equal in performance.

Stanine 5 includes scores 0.25 standard deviations above and below the mean (thus z scores of $+0.25$ to -0.25). Each stanine above or below stanine 5 includes another half of the standard deviation of scores. Again, Figure 4.1 shows the percentage of scores within each stanine.

Percentages Much of the accountability data required of schools today is reported as **percentages**. A percentage is calculated by dividing the number of students or schools receiving a given score by the total number of students or schools and multiplying by 100. Although percentages are useful measurements, some types of data may be distorted when percentages are used. When one examines changes over time, the result may change substantially, based on the size of the group being measured. No Child Left Behind requires that schools report the percentage of children from each ethnic group who pass the state test. If a small rural district had a small population of one ethnic group, their results could change substantially if only a few families move into their district. For example, one of the authors of this book worked with two districts with different student populations. District A, a rural district, had only two African American students in grade 4, and district B had more than 100 African American students in grade 4. Both of the students in district A failed the state test, resulting in a 100% failure rate. In district B, 50 students failed, resulting in a 50% failure rate. So based on percentage scores, it would seem that district B is doing better, even though they have many more students failing. The overall lesson here is that one should always examine both the total number and the percentage of persons when comparing measures.

Grade-Equivalent Scores A **grade-equivalent (GE) score** is a score reported in years and months for norm-referenced tests. So a GE of 3.4 means third grade, fourth month. A grade-equivalent score reports the grade placement for which that score would be considered average. This means that score provides an estimation of the performance that an average student, at a particular grade level and month, would likely demonstrate on this test. For example, Sasha, a first grader, scores a GE of 3.4 on an English Language Arts standardized test. This means that Sasha has performed the way an average third grader, in the fourth month of the academic year, would likely perform on this first grade test. It does not mean that the student is capable of doing work at this grade level. Remember, the test does not contain third grade content! It is a test for first graders. Sounds confusing? Absolutely! Although GEs are still used by some school districts, they are not particularly informative and are often subject to a great deal of confusion.

EVALUATING THE QUALITY OF STANDARDIZED INSTRUMENTS: RELIABILITY AND VALIDITY

Reliability and **validity** are the two criteria used to judge the quality of all standardized quantitative measures. *Reliability* refers to the consistency of scores, that is, an instrument's ability to produce "approximately" the same score for an individual over repeated testing or across different raters. For example, if we were to measure your intelligence quotient (IQ), and you scored a 120, a reliable IQ test would produce a similar score if we measured your IQ again with the same test. Reliability is a concept that is expressed in terms of correlations that are summarized in a **reliability coefficient**. The reliability coefficient can assume values from zero to +1.00. The closer to +1.00 the reliability coefficient is, the more highly reliable the instrument.

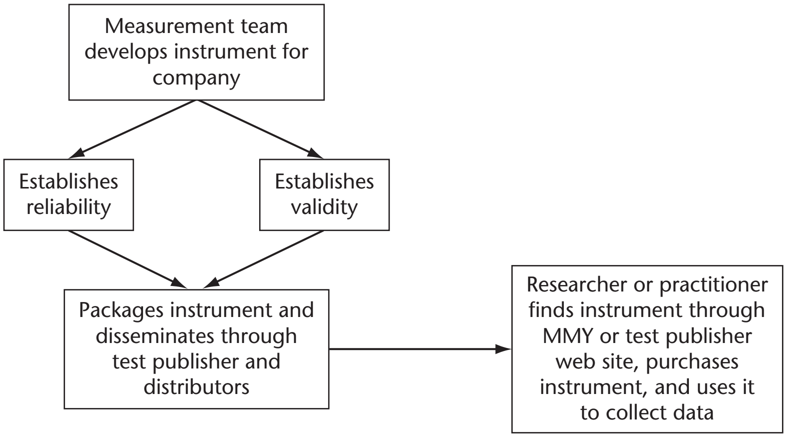
Validity, on the other hand, focuses on ensuring that what the instrument "claims" to measure is truly what it is measuring. In other words, validity indicates the instrument's accuracy. Take, for example, a fourth grade state math assessment that requires students to read extensive word problems in answering math questions. Some math teachers might argue that the assessment is not valid because students could know all the mathematical functions to answer the problems correctly but might be unable to answer the questions correctly because of low reading ability. In other words, the assessment is really more of an assessment of students' reading skills than it is of math ability.

Validity and reliability are typically established by a team of experts as part of the process of developing a standardized instrument. As part of the development process, the measurement team works to establish sound reliability and validity for the instrument before it is packaged and sold to researchers and practitioners (see Figure 4.2).

If an instrument does not have sound reliability and validity, the instrument is of no value. Therefore, it is important that beginning researchers and educators in general have some understanding of issues surrounding reliability and validity so that they are able to select the most appropriate and accurate instruments as measurement tools for their study.

In our initial explanation of reliability, we referred to *consistency* as meaning obtaining approximately the same score for an individual over repeated testing. Even the most reliable test would not produce the exact same score if given twice to the same individual. This difference is referred to as the **measurement error**. Error in measurement is due to temporary factors

FIGURE 4.2 Overview of the Process in Developing Standardized Instruments



that affect a person's score on any given instrument. Factors that influence reliability include:

- Test taker's personal characteristics (for example, motivation, health, mood, fatigue)
- Variations in test setting (for example, differences in the physical characteristics of the room)
- Variations in the administration and scoring of the test
- Variation in participant responses due to guessing

Because there is always some error associated with measurement, observed scores (the actual scores the individuals received) are made up of two components: the **true score** (the score representing students' real knowledge or abilities) plus some error. If a test were perfectly reliable, the true score would be equal to the observed score. Students and researchers alike understand that their score on a given test may not represent their true score. In essence, this means that unless an instrument is perfectly reliable, you can never know a person's true score. However, the simple calculation of the **standard error of measurement** (SEM) allows one to estimate what the true score is likely to be. The SEM takes into account both the reliability coefficient of the test and the variability of the scores of the norm group as calculated by the standard deviation. The formula for SEM is as follows:

$$\mathrm{SEM} = \mathrm{SD}\sqrt{1 - r}$$

where SD is the standard deviation of the test scores obtained by the norm group, and *r* indicates the reliability coefficient of the test calculated from a pilot group.

For example, consider the following situation. Test A has a reliability coefficient of 0.91. This suggests that the test is highly reliable, and therefore, a person's observed score is a relatively good estimate of his or her true score. The scores for test A produce a standard deviation of 9. Let us say that a student's observed score is 70. What would the SEM be for this example? Using our formula, it would be calculated as follows:

$$\text{SEM} = \text{SD}\sqrt{1-r} = 9\sqrt{1-.91} = 9\sqrt{.09} = 9 \times .3 = 2.7$$

The SEM for this example is 2.7. This means that the student who received a score of 70 might score 2.7 points higher (72.7) or lower (67.3) if she or he retook the test. All standardized tests should report the SEM. Knowing the SEM is important when a measure is used to make decisions about students (for example, placement or admission to programs). Decisions about students scoring near a cutoff point should be made cautiously, because retesting might result in the student scoring differently. Parents or teachers might consider asking that the student be retested or that additional information be considered.

Presented next are several types of reliability. Keep in mind that not all of these types have to be conducted on a single instrument. Typically, at least one is used to establish reliability during an instrument's design and development.

Stability or Test-Retest Reliability (Consistency Across Time) One form of reliability is called **stability** or **test-retest**. The purpose of stability is to show that an instrument can obtain the same score for an individual if that person takes the instrument more than once. "What is the importance of this?" you might be asking yourself. Well, imagine an intelligence test that gives a different score for an individual each time it is administered. The first time a student takes it, he receives a score of 140, indicating that he is a genius. However, the second time the instrument is administered, the student receives a score of 104, which equates to about an average intelligence. Which one is the correct score? How would a school psychologist or a researcher who is using this instrument interpret this information? The answer is that they couldn't, because the test is unreliable.

Stability of an instrument is also critical when one is trying to infer causality or that a change occurred in a study. Imagine a math assessment that was so unstable that a student received a score of 70 out of a possible 100 the first time the test was taken and a score of 90 out of a possible 100 the second time, even when no type of intervention or independent variable took place between the times of testing! Knowing the instrument's instability, how could you use the measure in a study to determine whether changes in student achievement in mathematics were the result of a new math program you planned to implement? You would never really know whether the change occurred because of the intervention, the unstable instrument, or both. Therefore, a researcher or a practitioner would want to avoid using an instrument that had poor test-retest reliability.

To establish stability, first, a pilot sample of persons is selected. Pilot samples in instrument development are usually large populations, which allow for the greatest degree of generalizability. Once the pilot sample has been established, participants are given the instrument, and scores are obtained for every individual. Next, some time is allowed to pass. The amount of time varies based on the type of instrument that is being developed. The most important thing is that enough time passes so that participants are unable to recall items and perform better on the instrument because of the first administration. An average wait time of 4 to 6 weeks between instrument administrations is usual. Following the second administration of the same test, the two scores are compared for each individual. Correlations are then calculated on the two sets of scores, which produces a correlation coefficient. Depending on the instrument's purpose, correlation coefficients ranging from 0.84 or stronger would be considered very strong; however, correlations from 0.35 to 0.64 are more typical and are considered to be moderately strong and certainly acceptable. Coefficients of 0.24 to 0.34 are considered to show a slight relationship and may be acceptable levels for instruments that are considered exploratory.

Equivalent-Form Reliability (Consistency Across Different Forms) Another type of reliability the measurement team might decide to establish for the instrument is **equivalent-form reliability**. This type of reliability is also referred to as *alternative form*. For this type of reliability, two forms of the same tests are given to participants, and the scores for each are later correlated to show consistency across the two forms. Although they contain different questions, these two forms assess the same content and knowledge, and the norm groups have the same means and standard deviations.

Internal Consistency Reliability (Consistency Within the Instrument) **Internal consistency** refers to consistency within the instrument. The question addressed is whether the measure is consistently measuring the same trait or ability across all items on the test. The most common method to assess internal consistency is through **split-half reliability**. This is a form of reliability used when two forms of the same test are not available or it would be difficult to administer the same test twice to the sample group. In these cases, a split-half reliability would be conducted on the instrument. To assess split-half reliability, the instrument is given to a pilot sample, and following its administration, the items on the instrument are split in half! Yes, literally split in half—often, all the even- and all the odd-numbered items are grouped together, and a score for each half of the test is calculated. This produces two scores for every individual from the one test: one score for the even-numbered items, and one score for the odd-numbered items. The Spearman-Brown prophecy formula is then applied to the data to correct for the fact that the correlations for internal consistency are being calculated on the two halves of the instrument and not the whole original instrument. For the split-half approach to be appropriate, the instrument must be long enough to split in half and should measure a single trait or knowledge domain. In determining the acceptability for coefficients, the research team would use the ranges discussed earlier in test-retest reliability.

Another approach used to establish internal consistency of a measure is to examine the correlations between each item and the overall score on the instrument. The question addressed here is whether a given score consistently measures the same amount of knowledge on the test. For example, consider scores on a 50-item history test taken by high school students. If all items are worth one point, students might get 40 out of 50 points by missing the first 10 items, the last 10 items, or some other mixture of 10 items. Can you say that all of these scores represent the same amount of knowledge? This type of internal consistency is assessed using either the Kuder-Richardson or Cronbach's coefficient *alpha* formula and is often referred to simply as internal consistency.

Validity

Along with the need to establish reliability for an instrument, the measurement team also must address the issue of **validity** by continually asking themselves the question, "Does the instrument measure what it is designed to measure, or is it measuring something entirely different?" To add to the confusion, an instrument can be extremely reliable but not valid in that it consistently obtains the same score for an individual but is not measuring what it intends to measure. For example, an educational research professor gives a midterm exam that she believes tests the content of the first half of her course. However, she inadvertently includes items from inferential statistics, which is covered in the second half of the course. This test would lack validity but would likely be reliable. That is, student performance would likely be fairly consistent over time. However, if you have a valid test that measures what it was intended to measure, the instrument also has reliability. So when constructing a test or using a standardized instrument, validity is the single most important characteristic.

As with reliability, validity is established during the piloting of an instrument. In most cases, it is done simultaneously with procedures to establish reliability. Validity is usually established through an in-depth review of the instrument, including an examination of the instrument's items to be certain that they are accurately measuring the content or objectives being tested and by relating scores on the instrument to other measures.

Content Validity **Content validity** examines the degree to which an instrument measures the intended content area covered by the measure (for example, curriculum, a personality trait, a set of behaviors). Let's assume one wants to examine the content validity of a test. In order to establish content validity, experts (including the classroom teacher who constructs the test) examine the objectives of the course or unit of study to determine whether all the areas of the content have been adequately covered on the test. In other words, does the test cover the breadth of the objectives covered by the teacher? This process involves an examination of the objectives and a comparison of the objectives and the items for a match between the two. In addition, the experts examine the amount of time devoted to each objective and the number of items represented by the test. For example, let's say an educational research professor is constructing a midterm. During the first half of the course, this professor spent one hour discussing sampling. The midterm, designed to cover all objectives of the first half of the course, is made up of 50 items. The test includes 30 items on sampling, leaving only 20 items for all of the rest of the objectives. Would this test have content validity? Obviously, this professor needs some tutoring on test construction.

Content validity also examines each individual item for its appropriateness. Does the item actually measure the knowledge or skills within the content area? Let's say a review of the educational research midterm discussed previously includes an item that has not been covered yet but will be covered during the second half of the semester. Having an item outside of the content area reduces the content validity of the test. Haven't you ever taken an exam where you come across an item that was never discussed in the class or in the textbook? If you have, this test probably lacked good content validity.

There may be times when an item is written to assess a particular concept, but it measures something else. For example, a math teacher might write an item to test student skills in single-digit adding and subtracting. However, the item was embedded in a word problem that used language beyond the students' reading level. Students who knew the math concept might still not be able to show their understanding of it.

Criterion-Related Validity Essentially, **criterion-related validity** involves the examination of a test and its relationship with a second measure. It reflects the degree to which two scores on two different measures are correlated. There are two forms of criterion-related validity. **Concurrent validity** examines the degree to which one test correlates with another taken in the same time frame. When might this be done? Let us say that you have developed a new instrument that is quicker and faster for measuring intelligence. To establish its validity, you might administer your new test, with an existing, already validated intelligence test, to a pilot group. You would then correlate the two sets of scores from your pilot group, and if you found a high correlation between the two tests, you could then offer your "new fast-and-easy intelligence test" as a replacement for the existing intelligence test, saying that it has evidence of concurrent validity.

Another type of criterion-related validity is **predictive validity**. Predictive validity uses the score on a test to predict performance in a future situation. In this situation, a test is administered and scored. Let us use the American College Test (ACT) that is taken by many high school juniors and seniors as an example. The ACT (the predictor) is used by colleges and universities around the country as a predictor of college success (the criterion). To determine its predictive validity, one would administer the ACT to a group of high school students and then allow some time to pass before measuring the criterion (college success—in this case, defined as one's freshman grade point average). Once the first-year grade point averages are released, they can be correlated with the ACT scores to determine if ACT is a successful predictor. If the correlation is high between the ACT and freshman grade point average (and there are some studies that suggest that this is the case), then the ACT is considered to have criterion-related validity in predicting college success.

Construct Validity **Construct validity** involves a search for evidence that an instrument is accurately measuring an abstract trait or ability. Remember that constructs are traits that are derived from variables and are nonobservable. They include complex traits such as integrity, intelligence, self-esteem, and creativity. In constructing an instrument to measure one of these complex constructs, one must first have well-developed ideas or a comprehensive theory about what the construct is and how people who vary in that construct would differ in their behaviors, abilities, feelings, or attitudes. Consider the construct of self-esteem. The Pictorial Scale of Perceived Competence for Children and Adolescents (Harter & Pike, 1984) grew out of their research on the development of self-esteem. They conceptualized that self-esteem for children comprised four components: cognitive competence, peer acceptance, physical competence, and maternal acceptance. They believed that children have an overall sense of self-esteem that is composed of these separate components. Questions on the instrument reflect this overall theory of what self-esteem is. Examining the construct validity of this instrument means examining whether it really captures the full meaning of the self-esteem construct.

Construct validity is one of the most complex types of validity, in part because it is a composite of multiple validity approaches that are occurring simultaneously. This might include aspects of content, concurrent, and predictive validity. Therefore, many researchers consider construct validity to be a superordinate or overarching type of validity. Fully establishing construct validity involves a lengthy process of collecting evidence from many studies over time. Following are some of the questions and procedures that might be involved in establishing construct validity.

- *Does the measure clearly define the meaning of the construct?* Evidence cited as support might include constructing a measure from a well-developed theory or examining whether its questions and parts correspond to research or theories related to the construct.
- *Can the measure be analyzed into component parts or processes that are appropriate to the construct?* In a test of creativity, the mental processes required by the tasks presented might be analyzed to see if they truly require creative thinking and not memory or reasoning abilities. Factor analysis is a statistical procedure used to see if clusters of questions identified in the statistical analysis match the groupings of questions identified by the test maker as the different parts of the test. So, for Harter and Pike's scale (1984), one would expect the items on physical competence to be clustered together by the factor analysis.
- *Is the instrument related to other measures of similar constructs, and is it not related to instruments measuring things that are different?* Researchers want to be certain that the instrument actually measures the construct and not something else. For example, let us consider the construct self-esteem. To validate their measure of self-esteem, Harter and Pike (1984) would correlate the self-esteem test with measures of similar constructs as identified in the literature. For example, if the literature says that self-esteem is related to body image and self-awareness, measures of these constructs should correlate with the self-esteem instrument. These correlations would produce confirming or **convergent evidence**. However, they would also want to make sure that the measure did not correlate with measures of constructs that differ from self-esteem. So scores on the self-esteem measure might be correlated with scores on measures thought to be unrelated or only weakly related to self-esteem (for example, attentiveness, emotionality, or physical activity level) to establish disconfirming evidence (constructs for which there was little or no correlation).
- *Can the measure discriminate between (or identify separately) groups that are known to differ?* Researchers will sometimes see if the measure yields different scores for

two groups who are expected to differ in the construct. Harter and Pike (1984) demonstrated that the self-esteem measure did distinguish between children who had been held back in school and those who were not held back.

Perhaps you can see why it takes a long time to establish construct validity.

Relationship Between Reliability and Validity

As you can see, both reliability and validity are critical to standardized tests. These concepts have an interesting relationship. A valid test is usually reliable. Yet, a reliable test is not always valid. Take the following case. A teacher has just covered a unit on the fall of the Roman Empire. She mistakenly includes many items on the test that were not included in her unit, thereby compromising the validity of the instrument. Remember the test was supposed to cover the fall of the Roman Empire. Let's say she gives the test to her students and they perform poorly. She decides to administer the test again the next week, and the students perform in a consistent way, giving the test reliability. However, the test would still not be valid! It did not measure what it was intended to measure.

It should be noted here that in order for a standardized test to maintain both its validity and reliability, it should be administered and scored as instructed in the test manual provided. Under most circumstances, failure to do so will compromise the quality of the instrument.

ISSUES IN FINDING AND USING STANDARDIZED INSTRUMENTS

You may feel like you've learned more than you ever wanted to know about standardized instruments. However, there are a few issues you should consider as you select and use any measure. These issues include testing students with disabilities, bias in testing, and finding the best instruments.

Testing Students With Disabilities

There may be some circumstances for which standardized procedures may have to be modified. You might think that this is a contradiction to the previous paragraph that suggests that deviation from standardization could compromise the validity and reliability of the test. While this is true under most circumstances, testing students with disabilities may require **assessment accommodations**. Assessment accommodations are changes in the standardized procedures. These accommodations are used to minimize or eliminate the potential impact of the student's disability that is not related to what is being measured. For example, a written standardized science test is administered to a class in which one student has cerebral palsy and lacks enough control to write clearly. A possible accommodation for this student is that the test be administered orally so that the student speaks his responses rather than writes. Since the test measures science and not writing ability, the reliability and validity of the test have not been comprised. It is important to note that the accommodation should not give the individual with a disability an unfair advantage over individuals without disabilities.

Many assessment accommodations are used to meet the needs of students with a variety of disabilities. Obviously the type of accommodation is dictated by not only the disability but the content area being tested. Assessment accommodations can include, but are not limited to, the following examples:

- Extended examination time
- Alternative settings (for example, study rooms, individual administrations, and so on)
- Changes in the way the test is administered (for example, the use of sign language, Braille or large-print formats, computer administered tests, and so on)
- Alternatives in the way responses are accepted (for example, oral exams, the use of a Braillewriter, voice-activated software, and so on)
- Alternative or modified tests (especially for students with severe disabilities)
- Adaptive devices and/or aids

Selection of the accommodation assessment should be done carefully and should always be designed to meet the individual needs of the students. Since these needs may change over time, assessment accommodations are reviewed and modified as appropriate.

Bias in Testing

Bias in testing received much attention in the 1970s and 1980s, when the results of standardized tests revealed gaps in the performance between members of different socioeconomic groups and racial and ethnic groups. Confronted with these gaps in performance, researchers sought to explain why these differences existed. This research produced two ways of identifying bias in tests. In the first, a bias review panel would examine these items keeping in mind the following questions:

- Does the test content presume knowledge or skills that are not shared by all social, ethnic, cultural, geographic, or ability subgroups?

- Does the test assume that all subgroups have had the same experiences?
- Does the test content offend any subgroups?
- Does the test use words or formats that favor one subgroup over another?

Any items that seem culturally inappropriate would be rewritten or eliminated. Examine the following test items and explain how they show bias.

- A question similar to this appeared on a fifth-grade math assessment: John goes into a department store with $40.00. A pair of slacks costs $17.99 with tax included. At the register the clerk indicates that the slacks are discounted by 10%. How much money does John have left when he leaves the store?

The bias in this item includes the following: student experiences with department stores, student understanding a pair of slacks is one item and not two; student not knowing the term *slacks*; student understanding discounted items.

- On a fourth-grade ELA assessment, the students were asked to identify the main idea of the following narrative: The Jones family is going on a ski vacation to Vail. The children are planning to use their new ski equipment, which includes boots, bindings, and new ski poles. The family has rented a chalet on the ski slope so that they can spend as much time skiing as possible. The children hope to be able to ski on the black diamond trail. Mrs. Jones is concerned about the children skiing on the black diamond trail. She has talked to her husband about having the children take lessons before they go up the slope. Explain why the children might be excited about this trip and why Mrs. Jones might be concerned for her children.

The bias in this item includes the following: student geographic knowledge about skiing, student understanding of black diamond trails, student familiarity with Vail, student understanding of terms like *chalet*, *bindings*, and *slope*.

The second approach focuses attention on the examination of specific test items to determine whether they produce differences in performance by different subgroups. It is not enough to identify items that result in different levels of performance for different subgroups, because these differences could reflect instructional differences rather than bias. Instead, the bias review panel compares subgroup performance on both individual items and the whole test. According to Schellenberg, "The group's performance on the item must be either better or worse than the group's performance on the test as a whole for the item to be eliminated" (2004, p. 8).

In addition to examining test items for bias, one should consider possible sources of bias in how the test is administered. Extensive research on this issue has shown that it is not essential for test examiners to be the same race, ethnicity, or gender as the test takers. However, having a test examiner who is "like the test taker" may increase the comfort of the person, especially if the test is administered individually. It is essential that test examiners not convey uncaring or prejudicial attitudes through their words or actions. For test examiners who differ in background from the test takers, it is recommended that the examiners:

- Understand possible cultural differences in the verbal and nonverbal behavior of the persons being examined
- Be attentive to student reactions and how these might affect taking the test
- Structure situations in ways that increase motivation to do well on a test, such as emphasizing the importance of doing well
- Be flexible in designing test settings that are comfortable for the person tested
- Be honest about the test and supportive of the persons tested

Another consideration in eliminating test bias is the need for all students to have the same degree of "test-wiseness." Students should be familiar with test formats prior to taking a test and may benefit from practice in taking tests. Students from affluent districts are often more likely to be offered opportunities to engage in test preparation in and outside of the classroom.

When interpreting scores from a norm-referenced standardized test, one must also consider what subgroups were included in the norm group. In other words, is the norm group representative of the group you are testing? In order to use the test information to determine how one's students are performing, there must be an appropriate norm group for comparison. Many tests now include district, local, state, and national norms and provide data on performance of various subgroups.

Finding Standardized Instruments

One source that many researchers use is the Mental Measurement Yearbook (MMY). University and college libraries and many large public libraries carry this reference book in their bound collections and through their electronic databases. Essentially, MMY is an overview of a wide variety of measurement instruments. Each instrument described in the MMY includes the following information:

- Complete name of the instrument
- The publisher, date of publication, and most recent edition
- A description of what the test measures
- The population for which the test was designed

- Procedures for administration and scoring
- A critique of the instrument, including its validity and reliability

In most cases, the description provided by the MMY is detailed enough to provide the researcher with enough information to know whether the instrument might be appropriate for his or her study. The MMY also provides enough information to generate an instrument section for a research proposal. Although this general description of the instrument may suffice initially, it is strongly encouraged that researchers obtain a copy of the instrument and administrative materials for more in-depth review. Standardized instruments are usually purchased by the researcher and include materials on test administration and scoring, individual testing booklets, and answer sheets. Researchers must purchase original copies of the testing materials from the publisher. Photocopying materials to save costs is tempting but illegal because it is a direct violation of federal copyright law. In some cases, standardized instruments cannot be hand-scored by the researcher and must be sent back to the publisher for scoring.

Review copies can be obtained by contacting either the publisher or testing company directly. In addition, many institutions may have test libraries that contain a wide variety of tests and other related measurement tools available for review. Special permission may be required by a faculty member for students to check out these materials. A list of Web sites with information about available standardized measures appears in Box 4.3.

Criteria for Selecting Standardized Instruments

Just because you find a test with a name that sounds like what you are measuring does not necessarily mean that you can use the instrument for your study. Selecting an instrument that is reliable, valid, and appropriate for the population you are studying is not only important but vital to the overall success of the entire study. An instrument may have a high degree of reliability and validity, but for a variety of reasons, it may not be appropriate for the population that you are intending to study. The language used in the instrument may be at a reading level far above the group for your study. Or the test may presume knowledge or experiences that you know your group does not have. If the norm group cited by the instrument differs substantially from the group you intend to study, you should question whether the test will be valid for your study.

As part of the review of the literature, a researcher should examine past studies to see what specific instruments other researchers in similar areas have employed. In addition, the researcher would also want to look for any discussion

BOX 4.3
Web Sites with Information on Standardized Tests

- *ETS Test Collection:* http://www.ets.org/testcoll/index.html

  The Educational Testing Service (ETS) is the largest private educational testing and measurement organization in the world. This Web site includes resources on testing and assessment for educators, parents, students, researchers, and policymakers as well as a searchable directory of more than 20,000 tests.

- *Buros/ERIC Test Publisher Directory:* http://buros.unl.edu/buros/jsp/search.jsp

  This Web directory is part of the Educational Resources Information Center (ERIC) Clearinghouse on Assessment and Evaluation. It includes the names and addresses of more than 900 commercial test publishers, as well as information on more than 4,000 tests.

- *Association of Test Publishers*: http://www.testpublishers.org/memserv.htm

  This is a nonprofit organization whose members are providers of tests and assessment tools for educational assessment, screening for psychological disorders, professional certification and licensing, and other educational or clinical uses.

- Educational and Industrial Testing Service (EDITS): http://www.edits.net

  EDITS specializes in measures for career exploration and counseling and assessments of psychological attitudes and personality traits in children and adults that relate to school and work.

- Psychological Assessment Resources, Inc. (PARI): http://www.parinc.com

  PARI is a commercial publisher of tests and software for assessment of achievement, development, and learning outcomes; intellectual and cognitive abilities; health-related behaviors; personality and psychological problems; and neuropsychological functioning.

- *Pro-Ed, Inc.*: http://www.proedinc.com

  Pro-Ed, Inc., is a commercial publisher of tests in the areas of speech, language, and hearing; special education, rehabilitation, and gifted education; psychology and counseling; early childhood intervention; and occupational and physical therapy.

- Western Psychological Services: http://www.wpspublish.com

  Western Psychological Services is a commercial publisher of assessment tools in the areas of clinical and school psychology, neuropsychology, special education, family therapy, speech, language, hearing, substance abuse treatment, human resource development, and occupational therapy.

in past studies or literature reviews that indicate that the instruments are not appropriate or accurate measures and why.

Finally, do not settle for the first measure that you find. Using key search words to search the MMY database should produce several instruments that you can inspect for possible use.

SUMMARY

Educational measurement is critical to good research and practice. Measuring student learning outcomes and understanding the context of learning (for example, demographics) is important to researchers and to practitioners. These activities require the collection of appropriate data.

Archival data are data that have already been collected using standardized instruments. In educational research, these data would be housed at the school building or district level and would include data on student absenteeism, detention or suspension, standardized test scores, and the like.

In addition to archival data, researchers and practitioners use standards instruments as data collection tools. Standardized instruments refer to a category of measuring tools that have been developed and piloted usually by someone other than the researcher who is doing the current study. An important feature of standardized instruments is that they are often standardized measures. Standardized measures include a fixed set of questions, a framework, and procedures for test administration. They measure specific outcomes with results compared with a well-defined norm group that has been given the measure at a previous time during the instrument's development. Norm-referenced, criterion-referenced, and self-referenced tests are three types of standardized tests. In general, standardized instruments measure five broad areas: achievement, aptitude, personality, attitude or interest, and behaviors. Achievement tests measure what students have already learned in school, whereas aptitude tests are designed to measure not what a student knows but to predict what the student can do or how he or she will perform in the future. Personality tests differ from achievement and aptitude tests in that they measure self-perceptions or personal characteristics. Attitude and interest scales are other types of standardized measures that assess a person's attitude toward a topic or interest in a certain area. These measures typically gather self-reported information from participants.

Reliability and validity assist researchers in evaluating the quality of educational measures. Reliability refers to an instrument's ability to consistently produce the same results for an individual over time. Validity focuses on ensuring that what the instrument claims to measure is what it truly measures. Several types of reliability may be established for an instrument. These include test-retest reliability, equivalent-form reliability, and internal consistency. The standard error of measure is an estimate of the overall amount of error contained in an individual score. Four broad types of validity that are established for standardized measures are content, concurrent, predictive, and construct. When searching for standardized measures, the Mental Measurement Yearbook is one source that contains a wide variety of measurement instruments and details about their purpose, targeted populations, administration, and reliability and validity. If archival data are used, the researcher should try to ensure that they are valid by examining any template or forms used in the original collection of the data, as well as interviewing anyone who might have had a role in the initial data collection.

While standardization is important to maintain the validity and reliability of a standardized instrument, there are occasions where procedures are modified. This is the case in the assessment of students with disabilities. Assessment accommodations may be appropriate when testing students with disabilities. The goal of these accommodations is to meet the individual needs of students while maintaining the integrity of the test.

Much progress has occurred in the area of bias in testing. However, there are several important considerations that should be made when selecting, using, and interpreting standardized tests to be certain that appropriate tests are being used to assess all students.

Tests may be found in a variety of locations. Most researchers and practitioners consult the Mental Measurement Yearbook (which is available online). There are many Web sites on standardized testing available on the Internet. Care should be taken when selecting standardized tests to be certain that the right test is being used for the right purpose.

KEY CONCEPTS

achievement tests
aptitude tests
archival data
assessment accommodations
attitude or interest scale
behavior rating scales and checklists
bias in testing
concurrent validity
cognitive measures
construct validity
content validity
convergent evidence
criterion-referenced instruments
criterion-related validity
curriculum-based measures (CBM)
disaggregated
equivalent-form reliability
grade-equivalent (GE) score
internal consistency
measurement error
norm group
norm-referenced instruments
percentages
percentile ranks
personality tests
predictive validity
projective personality tests
psychometricians
raw scores

reliability
reliability coefficient
self-referenced instruments
split-half reliability
stability or test-retest reliability
standard error of measurement (SEM)
standard scores
standardized instruments
stanine scores
test battery
true score
validity
z score

DISCUSSION QUESTIONS OR ACTIVITIES

1. Discuss the advantages and disadvantages of using standardized instruments. What types of educational constructs do you believe can or cannot be measured accurately through standardized measures? Discuss the reasons for your beliefs.
2. Examine the school report cards for one school or district in your state, and print out examples of descriptive statistics used. Discuss what the graphs and numbers show, and identify some additional ways that descriptive statistics might be used to display the data.
3. Examine the evidence for reliability and validity available for the standardized assessments developed or used (or both) by your state department of education. Discuss whether the evidence for reliability and validity is strong enough to justify the use of these tests as measures of student progress.

SUGGESTED READINGS

Kohn, A. (2000). Burnt at the high stakes. *Journal of Teacher Education, 51*(4), 315–327.

Kurpius, S. E., & Stafford, M. E. (2005). *Testing and measurement: A user-friendly guide.* Thousand Oaks, CA: Sage.

Popham, J. (2003). *Test better, teach better: The instructional role of assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Salkind, N. (2005). Tests and measurement for people who (think they) hate tests and measurement. Thousand Oaks, CA: Sage.

CHAPTER FIVE

QUALITATIVE MEASURES AND PROCEDURES

CHAPTER OBJECTIVES

- Identify several types of qualitative data and key features of each type of data
- Describe how observation is used as a data collection tool and the varying degrees of researcher participation in this process
- Discuss the purpose of observational protocols in qualitative research and what should be included in these protocols
- Identify issues involved in conducting observations and carrying out fieldwork
- Identify ways in which interviews and questionnaires differ and the advantages and disadvantages of each
- Distinguish among structured, semi-structured and unstructured interviews and outline the general procedures for conducting an interview
- Create an interview protocol that includes all required elements and questions appropriate to qualitative research
- Identify types of documents and images produced by researchers or participants and how these can be used as data collection tools
- Describe how archival documents and artifacts can be used for data collection in qualitative research
- Identify types of electronic sources used in qualitative data collection and discuss the pros and cons of this type of data

CHARACTERISTICS OF QUALITATIVE MEASUREMENT

Qualitative measurement is used in professional qualitative studies, mixed-method research, program evaluation, and action research. Qualitative research is characterized by flexible, naturalistic methods of data collection and usually does not use standardized instruments as its major data source. Instead, qualitative researchers typically develop their own methods for recording data and collect several different types of data. Qualitative data are often gathered in the form of words, pictures, or both. The tools that are used produce data that allow for rich and thick descriptions of the phenomena being studied. Tools for qualitative measurement must be flexible enough to allow recording of data on complex areas such as the social context for a group's interactions, cultural beliefs and customs, personal interactions and learning processes, and multiple viewpoints. Ultimately the researcher is the primary measurement tool in qualitative research; all data are filtered through the researcher's eyes and ears. Therefore, qualitative measures usually include tools to record the subjective experiences of the researcher for analysis. Table 5.1 summarizes these characteristics of qualitative measurement.

To be as scientific and unbiased as possible, the researcher must be systematic in the data collection process and record the data with accuracy. Qualitative researchers use a variety of research tools, but they often prefer to use observations, conduct interviews, and conduct document analysis (for example, data from school or public records, documents, pictures, or artifacts). In addition, qualitative researchers sometimes produce or have participants produce documents, such as journals or diaries, or images, such as photographs or videotapes. Electronic sources such as discussion groups or blogs are gaining in popularity as a qualitative data source. This chapter focuses on how these tools are created and used to collect data. Methods used to evaluate the quality of the data collected are discussed in chapters 6 and 7.

TABLE 5.1 Characteristics of Qualitative Measurement

Uses flexible, naturalistic methods
Tools are usually developed by the researcher to fit the goals of the study
Data collected are in narrative (words) or image (pictures, maps) formats
Multiple tools are used to record different types of data
Measurement tools are included to record the subjective experiences of the researcher

Gaining Entry to a Site and Maintaining Good Field Relationships

Before collecting any data, the researcher must gain access to the site selected for study. This involves obtaining permission to conduct observations or interviews, and it also means developing relationships with persons based on trust and mutual respect. Typically, the researcher must identify the persons who serve as **gatekeepers**, persons with official or unofficial roles who manage access to people and places at the site. Gatekeepers might be persons in high-level administrative offices or staff in the main office who control entry to a building and who have detailed knowledge of the people and resources in the building. Gatekeepers can assist researchers by identifying people or places for study that best fit the goals of the research. The principal might tell you that there is no space for private interviews at the school, but the head secretary might know of offices that are available at certain times and people who would be willing to let you use them. The detailed everyday knowledge of some gatekeepers is a valuable resource in getting to know who, what, and where to focus in one's initial observations. It is wise to spend time nurturing honest and respectful relationships with gatekeepers and other participants at a site while one is developing plans and tools for initial observations.

After gaining acceptance in a field site, one must work to maintain good relationships with participants throughout the study. Chambliss and Schutt (2003) offer the following suggestions for successful field research to beginning qualitative researchers:

- *Have a simple, one-sentence explanation of your project.* "I want to learn how urban students use technology skills inside and outside of school."
- *Listen actively and show a genuine interest in what participants say and do.* Most people appreciate being listened to and taken seriously. Your payback will be better data.
- *Show up for as many different events and activities as you can.* This builds trust and gives you deeper understanding of the setting.
- *Pay attention to everything.* The things that seem most insignificant at first may provide the deepest insight.
- *Write everything down that day.* Field notes should be recorded as soon as possible. If you don't, you will forget much of the important detail.
- Always remember: "It's not about you, it's about them" (Chambliss & Schutt, 2003, p. 176). So tell their stories, not yours!

Use of Observations

Although observation is a large part of how we learn (Bandura, Grusec, & Menlove, 1966), observation as a tool of research requires systematic and careful examination of the phenomena being studied. Specifically, researchers who use observation must conduct their research in a way that results in accurate, unbiased, and richly detailed information. We casually observe people's fashions, the way they wear their hair, and how they act, but this does not constitute observation as a research tool. Observation as a research tool requires training in both what to observe and how to record the observations. An initial consideration for researchers is whether they will take part in the activities that they observe or simply observe others without interacting with them.

Degrees of Researcher Participation When conducting observations, the primary goal is to gather data that are accurate and naturalistic, and, to the greatest extent possible, that reflect the reality of the situation as the participants see it. This necessitates that you as the researcher become familiar with the setting and that the participants are comfortable with your presence. Before the start of the observation, you must decide to what degree you will allow yourself to be involved in the setting. You can choose among varying degrees of involvement. According to Gold (1958), whose classification of observation is often used to describe the degree of participation of the observer, you could be one of the following:

- *Complete participant.* As a **complete participant**, you become a member of the group, and no one in the group is aware of the fact that you are also an observer. While this might allow a true "insider's" view, this type of covert participation raises ethical concerns because, in essence, you are deceiving the participants. A classic study using this researcher role was the study of street corner life in an Italian slum by William Foote Whyte (1955). Whyte spent four years living with a family in Boston's North End and conducted his study by spending large amounts of time interacting with members of neighborhood gangs.
- *Participant as observer.* When acting as a **participant as observer** you are an active member of the group and actively participate in the group's activities and interactions, but each member of the group knows that you are also serving a research role as an observer. In essence, a collaborative relationship is developed between the observer and the participants. Although this removes the ethical concerns presented by being a complete observer, you may compromise the natural interaction of the group.
- *Observer as participant.* Choosing to be an **observer as participant** means that you will observe participants' activities as unobtrusively as possible and minimize

the number of interactions with them, although you maintain a visible presence in the setting. Interactions with participants may occur, but they are likely to be more formal and structured, such as through interviews or very brief informal interactions. Although you certainly have a connection to the group, you will not likely participate in the group's activities. A study by Kidder (1989) provides an example of this type of researcher role. Kidder spent a year observing in a classroom. He rarely took an active part in classroom activities and mostly sat at the back of the room taking notes. He did interact with students in a friendly manner when they initiated contact. However, he did not encourage further interaction and rarely initiated communications with students.

- *Complete observer.* As a **complete observer,** you record observations passively in as uninvolved and detached a manner as possible. You might observe from the back of a classroom, but you are not a member of the group and do not participate in the group's activities. The observations might even be recorded covertly from behind a one-way mirror or by using public spaces. Malesky and Ennis (2004) used this type of covert observation in their study of posts on a pro-pedophile Internet message board. They observed the interactions between individuals on the message board without posting any messages themselves in order to examine the cognitive distortions that supported the members' acceptance of child-adult sexual contact as permissible. However, this type of covert observation violates the requirement for informed consent and voluntary participation in research; the benefits and problems from its use must be carefully weighed by an internal review board.

When selecting a role as an observer, you must decide the extent to which you want your activities as a researcher to be overt or covert and the extent to which you will be a participant or observer. These decisions may influence how the participants behave, respond, and react. As a qualitative researcher, you need to recognize the influence this might have on the reality of the phenomena being investigated. It is also possible that your role as a researcher might change during the study. You might start out as an observer as participant but find yourself increasingly drawn into group activities. In this case, you should consider whether the goals of your research study are best served by changing or maintaining your selected role.

Key Features of Careful Observation After determining your role, you need to consider *what* you will observe. According to Goetz and LeCompte (1984) and many other qualitative researchers, careful observation should include at least the following key features:

- *An explanation of the physical setting.* This is an overall physical description of the space. For example, in a classroom, this description includes the number of desks, the teacher's work station, the number of students, whether or not there were computers and, if so, how many, and any other unique features the researcher feels should be noted.
- *A description of the participants in the setting.* Careful description of the participants includes not only who is in the setting but also why they might be there and what their roles might be. In addition, any relevant demographic information should be included.
- *Individual and group activities and group interactions.* The researcher should observe the activities the participants are engaging in. What is going on in the setting? Are there rules that are being followed? Special note should be made of the activities that will help to answer the subquestions.
- *Participant conversation and nonverbal communication.* Because qualitative data often include direct quotes, conversations should be observed in such a way as to note not only what is being said but also how it is being said.
- *Researcher behavior.* Because the researcher is part of the setting, careful attention must be paid to the influence the observer has on the behavior of the participants. Does the researcher's presence in any way influence what is occurring in the setting?

As you can see, careful qualitative observation is a very demanding process! However, remember that good qualitative research includes multiple observations over time. The focus of the observation might differ from week to week, and the researcher might use several different tools to record observations at different times. Next we examine the types of tools used by qualitative researchers.

Developing Observational Tools The categories of careful observation just listed are too broad to be of much help in focusing one's observations. The categories suggest that the type of data will be more complex, richer in detail, and less predictable than the behaviors and attitudes measured using checklists, rating scales, and the other quantitative tools discussed in Chapter Four. Therefore, many qualitative researchers develop tools called **observational protocols** that are specific to the topic of their study. A well-designed observational protocol includes a set of questions and a recording sheet.

Observational protocols generally identify important areas that the researcher must attend to in the observations and provide an organized space for writing down brief descriptions of conversations, interactions, and observed natural behaviors. Although a qualitative researcher does not know exactly what will occur in a given setting, the observational protocol helps guide the collection of data in a systematic and focused manner.

Observational protocols may take different forms and include varying degrees of structure. Exhibit 5.1 presents a sample observational protocol used in a case study of a student in a preschool program. At the top of the first page, most protocols include a header that provides space to record background information such as the date, setting, participants observed, and observer's name. Below the header, in ethnographic or case study research, brief phrases or questions are listed, identifying the focus of the observation, such as types of actions by teachers or students, features of the setting, or interactions. In a research proposal, students might include the subquestions for the study to identify categories or areas that will be observed, although these may change after students enter the field based on the initial observations. Remember that what you observe might differ for observations conducted at different times and new protocols might be developed to fit the goals of later observation.

EXHIBIT 5.1
Sample Observational Protocol and Recording Sheet

Date of observation:
Time of observation:
Setting:
Participants:
Observer:
Research Question: How does a cognitively advanced preschool student interact with peers and teachers in his preschool?
Subquestions:

- What types of interactions does Max have with other students in his preschool?
- When and how does Max interact with his teachers?

Max's Behaviors and Comments	Teacher's Behaviors and Comments	Peers' Behaviors and Comments
Observer's Reflections:		

The bottom part of observational protocols usually includes an organized space for recording the details of what happens in a particular setting on a specific day.

Most qualitative researchers write down their observations in the form of **field notes**, which are written descriptions of what the researcher observes in the field and his or her reactions and feelings. To control for observer bias, qualitative researchers usually record both descriptive field notes and reflective field notes. **Descriptive field notes** summarize what the researcher sees and hears in the setting; that is, they *describe* what happens in the setting during the observation in as much detail and with as few subjective comments as possible. Descriptive field notes might include:

- Detailed descriptions of what persons do and how they interact
- In-depth descriptions of the setting and materials used by the participants
- Verbatim conversations and direct quotes

Reflective field notes include descriptions of the observers' feelings and thoughts about what he or she is observing. These are often recorded as observer comments at the bottom of the recording sheet or as separate entries in a field notes log after the observations. Reflective field notes allow researchers to reflect on their own feelings, values, and thoughts in order to increase their awareness of how these might be influencing their observations.

The recording sheet space for writing field notes can be organized in many different ways. One simple method is to divide the sheet in two columns using the left-hand (and larger) side for descriptive field notes that are recorded chronologically as events happen and the right-hand side for reflective field notes. Observational protocols may be more structured, with multiple columns identifying the type of data to be recorded. Sometimes the recording sheet takes the format of a log. A **log** is a running record of events, activities, and interactions that the researcher records to document what occurred on a given day. Typically a log focuses on factual details, and the entries are brief. Often entries include a date and sometimes the time of day to allow a fine grained analysis of the sequence of events and activities. A log may have a very specific purpose, such as recording the types of strategies used in a one-to-one tutoring session with a child.

Many protocols provide space at the bottom of the page for recording reflective field notes.

Field notes are usually hand written onto the protocol's recording sheet and typed up later. Exhibit 5.2 presents an example of the field notes for one observation recorded using the observational protocol presented in Exhibit 5.1.

No matter which tools you use, the observations or "raw data" produced by your observations will form the basis for the results and conclusions you draw in the study. As such, your recording of this information must be detailed, precise, and accurate. You might want to try out developing and using an observational protocol by completing Activity 1 at the end of this chapter.

Conducting Observations Conducting good observational research takes time and practice. Before beginning, you should carefully prepare for your observations by reviewing the subquestions of your study to give you a clear idea of what behaviors and activities you will be looking for; they will focus your attention. Many qualitative researchers conduct their first sets of observations without recording data to get a feel for what happens in the setting and identify what interests them most. Based on these initial observations, subquestions might be added or revised. Following these initial observations, you would develop observational protocols and recording sheets appropriate to your subquestions. As you conduct your observations, keep the following in mind:

- *Keep your observations short.* As your skills improve, you can increase the length of time that you are observing.
- *Be alert to the behavior, conversations, and activities of the participants.* You will want to remember as much as possible from your observations. Making a mental note or jotting down notes will be helpful.
- *Concentrate on specifics.* Avoid being global in your observations. Look for examples of specific behaviors.
- *Remain as unobtrusive as possible to avoid changing the activities and behaviors of participants.* Even if you have chosen to be a participant-observer, you still want to observe the natural activities and processes at the site rather than initiating or leading activities.
- *Vary your observations.* Alternate between examining the broad picture of what is happening and focusing in on specific individuals or activities.

Use of Interviews

Most qualitative research includes interviews. The interview might be the major data collection tool of the study (particularly when the behavior of interest cannot be easily observed) or may be used to corroborate or verify observations. An interview is a "conversation with a purpose" (Rossman & Tallis, cited in Chambliss & Schutt, 2003, p. 177) conducted with a person or a group of persons.

EXHIBIT 5.2
Sample Field Notes Using an Observational Protocol

Date of observation: March 15, 2001
Time of observation: 8:50 AM–10:00 AM
Setting: XXX Preschool; City, State
Participants: Max (four-year-old boy), Max's mother, Ms. Smith (teacher), 3 four-year-old girls, 2 other four-year-old boys
Observer: XXX
Research Question: How does a cognitively advanced preschool student interact with peers and teachers in his preschool?
Subquestions:

- What types of interactions does Max have with other students in his preschool?
- When and how does Max interact with his teachers?

Max's Behaviors and Comments	Teacher's Behaviors and Comments	Peers' Behaviors and Comments
Max enters room with mother at 8:50. He says good-bye to mother and turns to play with blocks in the building center area. Max responds "Hi" without looking up. Max ignores Ms. Smith and continues playing with blocks. Max at first ignores Ms. Smith but then drops a block and walks quietly to the	Ms. Smith looks up and sees Max. She says, "Hi Max." Ms. Smith calls all students to the center of the room for the opening activity. Ms. Smith invites Max to join the opening activity. When he does not respond, she repeats this in a more directive fashion, "Max, come and join us." Teacher smiles & sings with children. She looks at Max	Three other children are already in the room, playing in other areas. Two others arrive in the next 5 minutes. All other students stop what they are doing and come to the center of the room. Other students watch Max or Ms. Smith. Two girls are giggling but the third looks disapproving.

center. He sits quietly during the song looking around the room. Max goes back to blocks, but two girls are already there. Max tries to join in their play by sitting down and picking up a block. Max responds to girl by growling and continues to try to pick up blocks. Max plays by himself with blocks and ignores the girls. He sings to himself. Max scowls at the girl and grabs the block from her. Max growls at the girls and gets angry. He jumps up and knocks down their tower, screaming at them, "You can't hog all the blocks!" Max growls at the teacher but goes to the time-out corner.	but says nothing. After the song she tells children they can go to interest centers. Teacher is talking with another group of children. The teacher scolds Max for knocking down the tower and screaming. She sends him to the time-out corner.	All students but Max sing the opening song. One girl tells Max, "We don't want you to play. You will just mess things up." Second girl says, "Okay, you can play but don't knock down our building!" Girls play with blocks and ignore Max. They build a tall tower. As the blocks grow scarce, one girl reaches for the same block as Max. The girl says, "Hey!" and tries to take the block back from Max. The girls scream back and run to the teacher. They complain that "Max won't play nice!"
Observer's Reflections: Max clearly has difficulty entering group activities. Even when he does, there is almost no interaction with classmates. He does not verbalize his needs but simply growls indicating his anger and frustration with the situation. The teacher seemed to expect such behavior from him and did not seek his explanation.		

In interviews, as in surveys, persons are presented with questions to answer. However, questions on surveys typically require only brief responses. Interviews can provide much more depth and explore more complex beliefs, knowledge, or experiences than can a survey. One advantage of surveys is that they can be administered anonymously to large numbers of persons. Most interviewers meet face to face with their interviewees, so even when confidentiality is promised, interviewees might be reluctant to reveal sensitive information. In addition, interviews take time to administer, and usually studies using interviews involve smaller samples. One major advantage of interviews over surveys is interviews' flexibility. Although most interviewers construct interview questions prior to the interview, there is plenty of opportunity to ask for clarification of answers or ask additional questions on unexpected issues that arise. Table 5.2 lists the advantages and disadvantages of surveys versus interviews. However, if your goal is to obtain rich, narrative data for a qualitative study, interviews are preferred over surveys.

Types of Interviews Many qualitative interviews are conducted **one-to-one** (one researcher to one participant); the interviewer lets participants express their thoughts regarding a topic or experience in their own words. Interviews are conducted individually when the researcher believes privacy is essential, when the topics probed are individualized for each person, or when the interviewer wants to explore each person's responses in depth. The goal of an individual interview

TABLE 5.2 Advantages and Disadvantages of Surveys Versus Interviews

SURVEYS		INTERVIEWS	
Advantages	**Disadvantages**	**Advantages**	**Disadvantages**
Can be administered to large group of persons in short period of time	Requires extensive planning and high cost for mailing and follow-up	Can focus on small group of major interest	Typically involve small samples
Anonymity promotes honest responding	Anonymity precludes follow-up; no chance to probe answers	Flexibility to modify or individualize questions and probe responses	Time consuming to administer
Responses are brief and easy to summarize	Responses are brief and provide little depth	Extensive data provide in-depth information on a small number of persons	Summarizing and analyzing data is time consuming and complex

might be to determine the participant's feelings, interpretations, or reactions to a personal set of events (often referred to as a *critical-incidents interview*). For example, one might interview administrators to identify the turning points or defining events in their lives that led them to seek careers in administration. Or the focus of an individual interview might be on allowing each person to describe his or her life experiences in as natural a way as possible (also known as *life histories*). Individual interviews are preferred when interviews focus on deeply private issues. For example, one-to-one interviews might be conducted with students who lived through the terrorist attacks on September 11, 2001, to capture their individual perspectives. In this same situation, the students could be interviewed as a group if the researcher believes that group sharing of experiences would produce better information.

Group interviews are used when the goal is to explore the perceptions or experiences of small groups of persons who have some common basis for responding. Group interviews are sometimes called **focus group interviews** because there is typically an identified topic that forms the focus of the interview. For example, one of the authors took part in group interviews with parents living in low-income, urban areas; the focus of the interviews was on the issues that the parents faced in trying to ensure that their children engaged in healthy amounts of physical exercise. With focus group interviews, the researcher collects data from multiple participants and observes and records the interactions and group dynamics that unfold. The interactions and group dynamics may enable participants to build on each other's comments, producing ideas or details that would not occur in individual interviews. However, a group interview might make some participants feel uncomfortable about disclosing sensitive personal experiences.

Group interviews also pose administrative challenges. One has to arrange a convenient time and place for all participants. Managing the interview can be complicated, because several participants may respond at once and their responses may lead into complex stories. It may be difficult to identify speakers from tape-recorded data. In addition, groups might include some quiet members who are reluctant to talk and talkative members who monopolize time. Many researchers use two interviewers in group interviews, with one researcher posing questions and interacting with participants and the other researcher taking notes on interactions, handling recording devices, and providing a second set of ears and eyes to observe participant comments and manage interactions. To achieve optimal interaction, groups of 7 to 10 participants are best.

One other consideration is the diversity of the group. Typically focus group interviews are conducted with persons who are expected to have similar experiences or a common referent. If a group is too diverse, discussion might be affected if participants feel that their experiences differ too much from those already discussed. Also, the language used by different groups might vary in ways that make discussion difficult. For example, if you are conducting a group interview about a new curriculum used at a school, it might be helpful to interview administrators, teachers, students, and parents in separate groups. Teachers might be reluctant to criticize an expensive new curriculum in front of the principal. In addition, students and parents might use different words to describe features of the curriculum rather than the more academic terms used by teachers. In a study conducted by one of the authors (D.S.), students used the term "sloppy copy" rather than "rough draft" in describing their writing process.

Once you have decided to conduct an interview, you must determine the amount of structure you want in the interview. That is, will you conduct a structured, semi-structured, or unstructured interview? A **structured interview** is one in which the researcher comes to the interview with a set of questions, does not deviate from those questions, and asks the same questions of all the participants. Generally, in qualitative research, most interviews have some degree of flexibility and the researcher conducts either a semi-structured or unstructured interview.

Semi-structured interviews are typically planned carefully before the interview is carried out. The researcher develops an **interview protocol** (see next section of the chapter) that includes a list of questions or topics to be addressed in the interviews with all participants. Like an observational protocol, the interview protocol helps guide the collection of data in a systematic and focused manner. The interview is only semi-structured in that the researcher can change the order of questions, omit questions, or vary the wording of the questions depending on what happens in the interview. The researcher might also add other questions during the interview to probe unexpected issues that emerge. The procedures used in the interview might also be semi-structured. For example, in a focus group interview, the researcher might ask all group participants to respond to a question before soliciting additional discussion and comments. Semi-structured interviews are clearly identified as interviews, with specific times, dates, and topics identified in advance.

Unstructured interviews are more conversation-like and allow for the greatest flexibility. The researchers may simply jot down a list of topics that they want to cover in the interview and a couple of open-end questions to start. The interviewee sets the direction of the interview and the researcher listens and responds in a conversational manner, asking additional questions to probe for more detail or redirect the flow of the interview to areas that have not been discussed. Unstructured interviews can also include brief, spontaneous conversations with persons in everyday settings. The researcher may not record notes until a later time. This can pose ethical concerns; researchers conducting unstructured interviews need to make clear that they are gathering data for their study during the conversations. Unstructured interviews demand high levels of skills from the interviewer; beginning qualitative researchers are advised to start with semi-structured interviews.

Interview Protocols As noted earlier, an important component of conducting a good structured or semi-structured interview is the construction of a written interview protocol. An interview protocol includes a header containing places to record the interviewer's name, date and location of the interview, and background information on the interviewee. (In some studies, background information is collected before the interview using a questionnaire.) The header includes a brief script that is read to the interviewee, explaining the purpose of the study and how the results will be used. The header includes a statement of confidentiality. Below the header, the preliminary questions to be used in the interview are listed. Because the procedures for conducting qualitative interviews are flexible, the questions serve as a starting point. A good interviewer uses the questions to begin the discussion and then asks additional questions based on the person's responses. The interview may look more like a conversation than an interview that uses set questions and responses. Therefore the difference between a semi-structured and unstructured interview often lies mostly in the degree of planning before the interview.

Exhibit 5.3 provides an example of the questions from an interview protocol. This protocol was developed by Spaulding and Lodico (2004) to study community interest in a non-tuition-based alternative school. Imagine the impossible task of trying to collect interview data from thirty community members without such a protocol. How would you remember all the questions? How could you be sure that you asked each person the same question in the same way? Varying a question by changing a few words can sometimes dramatically shift the intent of the question. Notice how the protocol helps to ensure a degree of standardization during the data collection process.

If multiple interviews are being conducted, the interview protocol may change over time, reflecting the emergent nature of qualitative research. Questions or topics might be added or changed before future interviews are conducted, based on the responses of early interviewees.

Types of Interview Questions In a good qualitative interview, the interviewee talks freely in a naturalistic way. Interview questions should stimulate the interviewee's recall and thinking without suggesting that particular responses are desired. Questions that lead to a simple "Yes" or "No" response or to answers of only a few words should be avoided. A general rule of thumb is to avoid questions

EXHIBIT 5.3
Sample Interview Protocol for Feasibility Study

Community Member Questions

1. If you were to describe your community to someone, how would you describe it?
2. How has the community you just described changed over time?
   Probes: How has it changed:
   Demographically? Economically?
   Religious affiliations?
   School districts or school buildings?
3. Describe the relationship between your community and your local school district or building. Please describe the:
   Strengths
   Barriers
   Areas in need of improvement
   Relationship with parents and parental involvement
4. What kind of response from the community do you think an alternative middle school that is non-tuition-based would have? (Describe the program model.)
5. Do you believe an alternative school such as this would serve the community in areas where the public school currently does not? Please explain.

beginning with "Do you . . . ?" Instead begin questions with phrases like, "Can you explain . . ." or "Tell me more about . . ." or "What types of . . . ?" Typically the interviewer does not need a large number of questions; each question should encourage rich and detailed responding. However, there are several different types of questions that can be used to provide organization to the interview and to encourage deep responding.

Most interviewers start with a broad question that invites the interviewee to talk in a general way about the topic and to determine what the initial focus will be. One type of question is called the **grand tour question** because it asks interviewees to answer as if they were introducing someone to their setting or life. Sample grand tour questions are

- Tell me about a typical day in your classroom (could be asked of students or teachers)?
- How would you describe this school to a new counselor (or teacher, school administrator, or school psychologist)?

Additional interview questions would be written to ensure that the subquestions for the study are fully addressed. Remember that the subquestions are *not* your interview questions, but they do identify your major interests. Your interview questions should sound like questions from a natural conversation; subquestions identify the goals for your research study. For example, if your subquestion is, "How do high school teachers enhance literacy skills in their teaching?" you might include interview questions such as:

- What types of literacy skills do you include in your lessons?
- How do you teach these skills?
- How are literacy skills included or not included in your grading?

As the interviews proceed, the interviewer may want to follow up on comments made by the participants with questions referred to as **probes**. A probe is a follow-up question that is asked to get clarification about a response or to seek elaboration and additional detail. Probes are used when interviewees give responses that are too brief or unclear. By asking additional questions, the interviewer avoids making assumptions about what the person meant and encourages her to "tell her story" in more detail.

For example, say that you are interviewing a child like Max (discussed in Exhibit 5.2), who has difficulty interacting with his peers. You might ask him a grand tour question like, "Tell me some of the things that happened to you in school today." If he responds by simply telling you of an activity he enjoyed, and you are trying to solicit some conversation about whether he interacted with his peers in the activity, you might use a probe like, "Was anyone else involved in the activity? What did they do or say? How did you feel about that?" Probes can be used to extend the topics or situations that interviewees discuss. For example, if Max focused only on happy events in his school day, you might ask, "Was there anything that happened today that did not make you happy?" Some interviewers use hypothetical questions to get interviewees to discuss situations that they seem to avoid. For example, if an administrator claims that behavioral problems are not an issue at her school, you might ask, "What would you do if graffiti started to appear on the walls of your school?" Probes cannot often be prepared before the interview because some probes depend on what the respondent says in answer to your questions. Other probes that focus on asking a person to expand on a brief response can be written prior to the interview. It is important to use probes to gather additional information while recognizing when the respondent does not want to go any further with his or her disclosure. In other words, know when to stop probing!

Table 5.3 lists examples of interview questions and probes that might be used in qualitative interviews.

TABLE 5.3 Types of Interview Questions and Probes

Type of Question	Example
Grand tour question: Asks the respondent to provide an overall description of a setting or a set of typical activities	"*Tell me about your program?* What would a student considering joining it need to know?" "*How would you describe* your job to a college student considering a career in this area?"
Example question: Asks the respondent to give an example of something	"*Can you give me an example of* something you liked or did not like in this orientation program?"
Native language question: Asks the respondent to identify words or phrases used by a group	"*What do students call* teachers who try too hard to be their friends?"
Hypothetical question: Asks what the respondent might do or what it might be like in a particular situation; usually begins with "What if" or "Suppose"	"*Suppose* it is your first day at this school. What would it be like?"
Devil's advocate question: Challenges the respondent to consider an opposing view	"Some people would say that teachers who lose their jobs did something to bring it about. *What would you say to them?*"
Ideal position question: Asks the respondent to describe an ideal situation	"*What do you think* the *ideal* administrator would be like?" "*What would your ideal* school look like?"
Interpretive: Asks the respondent to make a judgment	"*What do you think* are the strengths and weaknesses of this program?"

Source: Merriam (2009).

Conducting a Good Interview Although qualitative interviews are like conversations, conducting a good interview requires a high level of skill and extensive practice with those skills. One of the most important skills to master is **active listening,** which involves putting aside one's own presuppositions and listening carefully for the meanings being expressed by another. A good interviewer listens attentively and makes interviewees feel that their responses are important. This interest is communicated in many verbal and nonverbal ways, including quietly listening, nodding to acknowledge responses, allowing periods of silence for interviewees to think, asking questions that show the interviewer is paying attention, and taking and reviewing notes. Active listening also means that the interviewer is thinking about what is being said, what needs to be probed, and what other areas need to be addressed. If the study includes multiple interviews with participants, you can review your interview notes between sessions to determine issues that need further exploration. However, it is better to obtain rich and detailed information as the issues arise in the interview rather than repeat issues across interviews.

Good interviewers are acutely aware of their own behavior and avoid doing things that might change or bias what the interviewee says. If you nod your head only when the interviewee is critical of charter schools, you are not likely to hear positive comments about charter schools. If you notice a stiffening in response to a question, you might reorder your questions to address less sensitive issues first. If you laugh at a response and then observe that the interviewee stops talking, you need to consider whether your response was inappropriate. Mastery of all skills involved in interviewing takes time and practice, but fortunately it is an enjoyable task!

Before you begin the interview, it is essential to check your recording equipment and make sure that the location for the interview is quiet and private enough to enable the respondent to feel comfortable and relaxed. Ideally, the location should be selected for the convenience of the interviewee. During the week before the interview, it is a good idea to call and confirm the date, time and location.

Another consideration is how to dress for the interview. Should you wear a suit to appear professional, or would this distance you from respondents? While there is no hard and fast rule regarding dress, it is best to aim for a professional image appropriate to the setting. Your dress should not be extremely different from others in the setting, but you should also not insult participants by dressing down. A similar consideration is whether to change your style of communicating with respondents. Knowing some of the phrases used by a group may help you to communicate more effectively and is good preparation for interviewing. This knowledge is especially important if you are taking a participant-observer role in a setting. However, your use of the jargon or communication patterns of the group might come across as inauthentic or, worse, demeaning. Again it is best to be yourself and use a communication style that feels natural. A typical sequence of steps in conducting an interview follows.

1. *Begin the interview by reintroducing yourself.* Because you are doing qualitative research, you will likely have had some contact with the participant before the interviews. You may want to introduce the general topic that you will be discussing and the purpose of the study as outlined in your interview protocol.
2. *Remind the participant of the confidentiality of his or her responses.* At this stage, the participant and/or guardian may have already been told about confidentiality issues and agreed to participate in the study. However, it is important to review that information at the beginning of the actual interview.

3. *Obtain general descriptive information.* This type of descriptive information could include information about the participant or the issue or phenomenon being studied.
4. *Present your questions starting with the least sensitive or most general questions.* If you have not had extensive contact with the interviewee prior to the interview, the early part of the interview may be spent building trust and rapport. The interviewee may be trying to decide what it is safe to share with you. Chambliss and Schutt (2003) recommend that interviewees should be treated with respect, as research partners whose time is valuable. By showing genuine interest in responses and establishing a respectful and professional tone, you can gradually build trust and rapport that will permit you to probe responses more deeply or move on to more sensitive topics.
5. *Throughout the interview, strive for neutrality.* To maximize what participants tell you, it is particularly important that you are a good listener and nonjudgmental in your reactions. Be sensitive, and never act shocked or upset by what you hear. Being judgmental is likely to limit interaction and may cause the participant to question whether he or she should trust you. However, you must recognize that because you are conducting a qualitative interview, your values and personal biases, as well as those of the respondents, are a factor related to the kind of information you gather. As such, anticipate and document these to the best of your ability and consider the possible influence these might have on the data collected.
6. *Record the interview data.* There are many ways to record the information given by respondents during an interview. Many qualitative researchers prefer to tape record their conversations to preserve the integrity of the data. This is particularly important because many qualitative studies include verbatim responses as part of the data analysis. In addition, some researchers take field notes or jot down some key responses. Gray (2004) points out that note taking can help the interviewer formulate new questions or locate quotations in a tape transcript at a later time. The process of writing also buys time for you to consider what to ask next. Note taking also serves as a nonverbal message that the interviewee has said something important. If you choose to take notes, be sure to fill in additional information immediately after the interview.

Use of Documents, Images, and Artifacts

Documents, images, and **artifacts** are other types of data collection tools used in qualitative research. Documents are printed or written records that may have existed before the start of the study, such as a personal diary, or that were created after the study began, such as student essays. Images can include maps or diagrams of a classroom or program site or photographs or videotapes of events at a setting. Artifacts are objects used in the setting such as a map, textbook, or desk.

Documents and artifacts produced by the participants as part of their regular lives generally include familiar things like public records or reports, minutes from meetings, personal letters, bulletin boards, newspapers, yearbooks, or instructional materials. Typically these documents are collected from the site and their content analyzed. For example, Hoffman (2004) examined how student diversity was represented in high school yearbooks and compared yearbook content to information obtained through interviews with students. Some documents and artifacts, such as bulletin boards, meeting minutes, newspapers, or yearbooks may be publicly available documents. Other private sources such as letters might take considerable effort to collect. An example of creative use of personal documents occurred in a study by Hubbard (1989), who collected notes written by sixth grade students to each other to see how students used literacy skills to communicate in their "underground classroom culture."

At times, a researcher might ask participants to record their thoughts, feelings, or actions in a tool created for the study, such as a **journal** or a log. Logs are generally used when brief entries focused on descriptive details are desired, while journals typically are used to record reflections on feelings and thoughts as well as descriptive data. To structure the data generated in a journal, the researcher often creates a set of instructions regarding what to record, how often and what the expected length of each entry would be. In addition, a couple of broad questions might be posed to guide entries and an organized space provided similar to the protocols described earlier.

The decision to use documents, images, and/or artifacts is driven by the research questions you have asked. Again, turning our attention back to Max, say that you are conducting a case study of Max to determine the type of peer interactions he has with his classmates. What possible documents might you want to explore? First, there might be some school records of his attendance or any disciplinary action taken by the school director. Furthermore, you might want to ask Ms. Smith, Max's preschool teacher, whether she has kept a journal to record the behavior of the children. The records and journal could provide you with further insight into Max and his relationships with his peers.

Again returning to Max, assume that we want to examine his interactions with peers and adults *outside* of school. We might ask his mother to keep a log of his interactions with adults and peers at home and in the neighborhood. The entries could be very brief, identifying the date, the students he plays with, the length of time, and the type of activity. Space for recording additional thoughts about the interactions might also be provided. For Max's interactions in the classroom, let's assume that Ms. Smith did not keep a journal or that the director of the preschool does not keep a disciplinary log. As the researcher, you could ask

Max's mother to keep a journal of Max's responses to a set of questions about school asked of him after school. In this case, the journal would be a document created during and as part of the study. Some researchers prefer to use unstructured journals in which participants are free to record their ideas and feelings in any way that they wish. Entries might include drawings or diagrams to represent participants' understanding of a topic. For example, Shepardson and Britsch (2000) asked first- and second-grade children to keep journals of the materials and activities in a one-week Earth Science course. Children's journal entries included drawings that illustrated their understanding of the materials and science processes used in the class.

A researcher may create a journal of her or his own that may serve one of two purposes. First, the journal might be used to record and examine the researcher's subjective impressions during a study as a way to control researcher bias. Second, the journal might provide a flexible space for recording and analyzing some types of data. Journals might include drawings or diagrams of an educational setting as well as notes to the researchers (referred to as memos) to remind them about issues that they want to reexamine in future observations. The format for a researcher's journal is usually open-ended; the researcher can record any information, thoughts, or feelings he or she feels are important.

Maps and diagrams are tools that record the physical arrangements of research sites and how persons use that space. For example, a schematic map or drawing of Max's classroom might be created and used to record data regarding where he prefers to play. Does he spend more time in a solitary play area or one where there is the possibility of group interaction? Such a diagram could serve as a recording sheet for observations and might include notes on how the number of peers at different locations changes or remains the same as Max moves about the classroom.

By examining a map or diagram of a school or classroom without people present, a researcher might be able to better understand how certain values are reflected in the physical allotment of space. How much space in a school is given to technology versus physical education or the arts? Where are the spaces devoted to working with students with special needs? How is space in a classroom arranged to encourage group versus individual work?

Photographs and videotapes are often used to capture the rich detail present in natural settings and activities. Researchers might record their own images by videotaping activities occurring in a classroom or program. An alternative is to ask participants to record the images that they feel are important using cameras or videotapes. Images provide complex data that is open to many different interpretations and therefore vulnerable to researcher bias. Therefore, researchers often ask participants to discuss the images that they have recorded through either written comments or interviews. An example of the use of images as a tool in qualitative research occurred in a study by Garcia (2009), who documented youths' perceptions of their towns. Garcia asked the youth to photograph places that they liked or wanted to improve and then interviewed them about why the places were meaningful to them.

Use of Electronic Sources

Electronic sources are a relatively new type of data collection tool. Electronic sources can include both publicly available data, such as that available in blogs, discussion groups, social networking spaces, and Web sites, or private data such as e-mails. Many school organizations create Web sites with extensive information about school activities. For example, in a study of school-based ally groups affiliated with the Gay, Lesbian, Straight Education Network, Kennedy and Voegtle (2000) found extensive information about the groups' goals and activities in the mission statements, minutes of the groups' meetings, and discussion group threads available through their Web sites.

Electronic sources provide a wealth of information and are easily available; however, they do present ethical and research challenges. Since data from discussion groups and Web sites are publicly available, some researchers argue that informed consent is not necessary as long as confidentiality is protected. Others (Gray, 2004) argue that participants have a right to know that their information is being used in a research study. However, announcing that one is collecting data for a study may taint the subsequent entries. Gray suggests that, at a minimum, researchers should contact the moderator of a discussion group or the creator of a Web site to gain permission for the research. This is the procedure followed by Kennedy and Voegtle (2000).

Another question is whether it is okay to report information from a discussion group in an external publication such as a research report. Participants in a discussion group may well consider their disclosures as public only to other participants in the discussion group. There is also a danger that comments made in an electronic source may be taken out of context and thus misinterpreted. Although this is a common problem in collecting qualitative data, the brief and interactive nature of entries in electronic sources increases the danger of decontextualized representation.

Many electronic data sources contain information that is anonymous, and it is not possible to verify that participants are who they say they are. A researcher may not be certain about the gender, age, ethnicity, occupation, or location of participants. Even if this background information is available, it should not be reported because it might result in violations of confidentiality. Overall, we recommend that you think deeply about the ethical issues involved in using electronic sources and treat them with the same respect that you would give to participants that you meet face to face!

SAMPLING IN QUALITATIVE RESEARCH

Qualitative researchers use sampling to select a wide variety of data sources. People might be sampled for interviews or observations, and times or places for observations might also be sampled. Documents and artifacts, such as textbooks, student work, or issues of a school newspaper, might also be sampled. Qualitative researchers select their data sources based on the research questions being investigated. As you already know, the researcher's primary concern is to study individuals in their natural context, with little interest in generalizing the results beyond the participants in the study.

Types of Sampling

The sampling procedure most often used in qualitative research is **purposeful sampling**. According to Patton (1990), "The logic and power of purposeful sampling lies in selecting information-rich cases for study in-depth. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the research" (p. 169). So the goal of purposeful sampling is *not* to obtain a large and representative sample; the goal is to select persons, places, or things that can provide the richest and most detailed information to help us answer our research questions.

If a researcher is conducting an ethnographic or case study, she or he might identify **key informants**: persons who have some specific knowledge about the topic being investigated. For example, let us say that you are interested in discovering how an inner-city after-school program influences the lives of a group of youth of color. To get an insider's perspective, you would select a group of participants in a particular type of after-school program and who have in-depth knowledge of the program. You might further decide to sample after-school programs with a science focus aimed at elementary school students. This is a type of purposeful sampling referred to as **homogeneous sampling** because all participants share similar characteristics (for example, attending inner-city elementary schools with after-school science programs).

Qualitative researchers have identified many different types of purposeful sampling procedures. Miles and Huberman (1994) identify 17 variations and Patton (1990) identifies 15 strategies for purposeful sampling. Although all of these strategies are useful, Table 5.4 identifies the ones we believe to be most useful to beginning researchers. Following Creswell (2002), we have organized the strategies based on whether they are used before (the first six strategies) or after (the last three strategies) data collection has begun. If you are considering doing a qualitative study, consider which of these strategies you might use.

TABLE 5.4 Summary of Purposeful Sampling Strategies

Type of Purposeful Sampling	Purpose and Explanation	Example
Typical case sampling	Individuals are selected because they have characteristics or experiences that are representative of many others.	Teenage participants who report having at least one good friend are selected for a study of peer support.
Extreme case sampling	Individuals are selected because they represent extreme differences, for example, highest and lowest, best and worst in certain areas.	Students whose attendance rates are in the lowest and highest 10 percent of their class are selected to be interviewed about school belongingness.
Maximum variation sampling	Individuals are selected so that the sample includes the greatest possible diversity in the characteristics or experiences under study.	Participants are selected so that opinions ranging from pro-stem cell to anti-stem cell research are represented.
Critical case sampling	Individuals are selected because they can "make a point dramatically" (Patton, 1990, p. 174).	Individuals who had a loved one die in the Iraq war are interviewed regarding their positions on the war.
Homogeneous sampling	Individuals with similar attributes or experiences are selected for the sample.	Parents who live in rural areas and have children who attended preschool are selected.
Purposeful random sampling	One of the other purposeful sampling procedures is used to identify a group, and then individuals are randomly sampled from the group.	Critical case sampling is used to identify a group of individuals who lost a loved one in the Iraq war, and then twenty individuals are randomly selected to study.
Opportunistic sampling	Individuals who were not initially identified for the study are sampled because the opportunity presents itself.	In a study of how technology is used in art education, you discover that many of the teachers have contacted a local museum of media education. You decide to interview the director and educator from the museum as well as the teachers.
Snowball sampling	A few individuals who have certain characteristics or experiences are selected first. Then they are asked to nominate others with similar characteristics or experiences. The nominees may also be asked to identify additional potential participants.	Teachers who use music in their instruction are studied and asked to nominate colleagues who do the same. All teachers are included in the study.
Confirming/ disconfirming sampling	Individuals are selected in order to see if they support or disconfirm emerging explanations and theories.	In a study of allies to persons who are gay or lesbian, a campus activist who becomes engaged to a homophobic person is studied to see if her activist activities decrease.

If your study is a qualitative one, you will want to carefully consider the types of purposeful sampling discussed and which ones will best answer your research question.

Sampling and Collecting Data in Culturally Diverse Settings

Qualitative researchers often collect data on settings or groups that are different from themselves. While such research is motivated by a desire to understand these persons or settings, it raises questions about the researcher's ability to fully and sensitively understand and represent the setting and the experiences of the participants. Can a male researcher really understand the experiences of women who were sexually abused? Will the women be willing to share their stories with him? Can middle-class researchers appreciate the difficult realities of families that are homeless? Can a white researcher fully understand the day-to-day experiences with racism of African American youth? Can adult researchers fully understand the thinking of children who have grown up with technology as an ever-present and essential tool for communicating with peers?

Ideally, a research team should include partners who come from the setting or group under study. It is essential to communicate a genuine and respectful desire for key informants to educate you on cultural norms and issues and to feel free to correct or admonish you when your behavior or assumptions are incorrect or inappropriate. Tools for reflection on the researcher's biases and assumptions are vitally important, as are opportunities to discuss one's reflections with **peer debriefers** or key informants, especially persons who are members of the group being studied.

Researchers in culturally diverse settings should also attempt to gather culturally relevant information about all aspects of the setting or group before entering the field. Become familiar with the important events, foods, newspapers, people, and places in a community as defined by that group. If you are studying a Muslim neighborhood, find out what restaurants or events are listed in the local or community papers serving that neighborhood. Learn about the history of the neighborhood by reading local history books or talking with key informants. Find out about subgroups within the neighborhood and how they view each other. Don't make assumptions about surface similarities between groups. For example, Lodico and Voegtle (2005) learned that some Latino youth in one urban school chose different friends based on their specific ethnic group identification. Puerto Rican youth allied closely with African American students, while Mexican youth hung out more with Caucasian students.

It is helpful to learn about the conversational and interactional styles of the groups under study. Be aware of cultural norms regarding nonverbal behaviors such as comfortable distances for conversations, whether eye contact is maintained, whether criticisms or corrections are delivered directly or indirectly, and whether there are limits to interactions based on gender or religion. Some subgroups develop their own vocabulary or grammatical expressions. Gray (2004) recommends that one keep a list of terms and idiomatic expressions and how these are used. Of course, you should inquire about any unclear terms with your key informants. Some groups have verbal customs used to deliver information or instruction. For example, many Native American groups use storytelling to instruct children about socially acceptable behaviors or to keep them safe from hazardous situations (Bruchac & Bruchac, 1998). Novellas or mini-dramas are commonly used in Latino communities to communicate important health information (Quinn, Hauser, Bell-Ellison, Rodriguez, & Frias, 2006).

SUMMARY

Qualitative researchers typically develop their own flexible, naturalistic methods for collecting several different types of data and for analyzing the researcher's own potential biases. Methods of data collection in qualitative research typically include observations, interviews, documents and artifacts, and electronic sources that are either collected from the setting or created by participants at the request of the researcher.

Observations may involve different degrees of researcher involvement, including complete participant, participant as observer, observer as participant, and complete observer. Qualitative researchers must first identify gatekeepers—persons with an official or unofficial role who manage access to people and places at the site—and then maintain good relationships with participants in the setting. In their observations, qualitative researchers should note the characteristics of the physical setting in which the observation is taking place and the roles, interactions, and methods of communication of the individuals being observed. Observational protocols developed prior to the observation provide space to record both descriptive and reflective field notes. Qualitative subquestions are included on the protocols to guide the observations.

Interviews are another key method for data collection in qualitative research. Interviews can be conducted in small groups or individually with participants and the researcher. Group interviews are also referred to as "focus groups," in which the researcher has an opportunity to collect data from multiple participants and to record the interactions and group dynamics as they unfold. Interviews may be structured, semi-structured, or unstructured. Interview protocols should include a space to record background information, a script for describing the purpose of the interview, and the preliminary interview questions. A good interviewer should start out by reintroducing himself or herself, restating the purpose of the study, and reminding the participant about confidentiality of responses. Next, general demographic information should be obtained from the participant, and the interviewer should use protocol questions as a starting point and probes to gather more in-depth responses from participants where needed. During the interview the interviewer should practice active listening and strive for neutrality. When completed, the interview should look more like a conversation than an interview with set questions and responses.

Documents, images, and artifacts can also be used as sources for qualitative data. Documents created during the study include logs, journals, maps, or diagrams. Images may include photographs and video or digital recordings made by the researcher or participants.

Electronic sources, such as blogs, discussion groups, social network sites, and Web sites are increasingly used as sources of data although these present ethical and research problems.

Qualitative researchers use a variety of purposeful sampling strategies to select participants before and after data collection has begun. Researchers in culturally diverse settings may employ several different strategies to make sure that they understand and accurately represent the culture and experiences of the group.

KEY CONCEPTS

active listening
artifacts
complete observer
complete participant
descriptive field notes
documents
field notes
focus group interview
gatekeepers
grand tour question
homogeneous sampling
interview protocol
journal
key informant
log
observational protocol
observer as participant
one-to-one interviews
participant as observer
peer debriefer
probes
purposeful sampling
reflective field notes
semi-structured interview
structured interview
unstructured interview

DISCUSSION QUESTIONS OR ACTIVITIES

1. You are conducting a qualitative study on one of the following research questions. Discuss what data sources you might use and how you would go

about selecting them. Specifically, discuss whether you might use interviews, observations, documents, artifacts, images, or electronic sources and what type of purposeful sampling you would use to select persons or sources.

   a. How do students use literacy skills in their activities outside of school?
   b. What changes, if any, occur in preschool activities and interactions when the preschool begins to admit children with intellectual disabilities?
   c. What are the experiences of gay or lesbian parents in interactions with school staff and teachers?

2. Record an interview by a person whom you consider to be a professional interviewer. Think about what types of preparation the interviewer may have had to do prior to the interview. Analyze the types of questions used by the interviewer. How did the interviewer begin the interview? What areas did the interviewer probe? What additional questions might you have asked? Is there evidence of researcher bias in any part of the interview? If possible, have a friend listen to the same interview and answer these same questions. Compare your answers to see if you perceived the same or different things in the interview.
3. You are a researcher studying the use of assistive technology with a third grade student who is learning to read. The student was recently diagnosed with dyslexia. You want to gather data from a variety of sources to document this process. Document from whom you would gather data and develop the appropriate protocols.

SUGGESTED READINGS

Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative research for education* (3rd ed.). Boston: Allyn & Bacon.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2005). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Glesne, C. (2006). *Becoming qualitative researchers: An introduction*. Boston: Pearson Education.

Wolcott, H. F. (1995). *The art of fieldwork*. Walnut Creek, CA: AltaMira.

CHAPTER SIX

QUALITATIVE RESEARCH

CHAPTER OBJECTIVES

- Summarize the key characteristics of qualitative research
- Compare and contrast the different types of qualitative research, including narrative inquiry, phenomenological research, ethnography, and case studies
- Identify the major characteristics of qualitative research
- Discuss theoretical assumptions underlying different types of qualitative research
- Outline the basic steps in conducting qualitative research studies and discuss how completion of the steps might differ for different types of qualitative research
- Understand the criteria used to evaluate qualitative research and be able to use these to critique methods used in qualitative studies

RESEARCH VIGNETTE

David tutors students who are English language learners (ELL) in a large urban middle school. He has observed that teachers and school staff sometimes treat the students who are ELL and their parents differently than they do students for whom English is the native language. They use condescending or negative phrases when describing the families. Some of his students have shared stories with him about how teachers make assignments easier for them or assume that their parents are not able or willing to help them with homework. David wants to conduct a research study that will provide an in-depth look at the students' and parents' experiences. He decides that a qualitative study will best enable him to capture the rich details about the school experiences of the students and their families. His work with the students has shown him that these families are often pushed aside in the educational system, and he wants to ensure that his research will help to tell their stories in their own words. He decides to use narrative inquiry, a type of qualitative research that focuses on personal, firsthand accounts of people's lives. David believes that the detailed, personal stories created through narrative inquiry will provide the most powerful research evidence of inequities in the treatment of students who are ELL. He selects three families and conducts extensive interviews with the students and parents about their experiences from the time that they entered the United States to the present. He works closely with the families to select and organize data generated by the interviews. In the final research report, he briefly explains his research methodologies and provides social and cultural background on the school; however, most of his report consists of the stories of the families told in their own words. As the researcher, David "restoried" their stories, organizing the events and interpreting them.

UNDERSTANDING QUALITATIVE RESEARCH

Qualitative research, also called *interpretive research* or *field research*, uses methodologies that have been borrowed from disciplines like sociology and anthropology and adapted to educational settings. Qualitative researchers, as you already have learned, use the inductive method of reasoning and strongly believe that there are multiple perspectives to be uncovered in their research.

Qualitative researchers focus on the study of social phenomena and on giving voice to the feelings and perceptions of the participants under study. This is based on the belief that knowledge is derived from the social setting and that understanding social knowledge is a legitimate scientific process. David's choice of a qualitative approach to study the experiences of students who are ELL illustrates how important it is to examine complex social situations in depth and to focus on persons who are often silenced or overlooked.

Although there are many types of qualitative research, we discuss in this chapter the methods that are most commonly read and utilized by our students, most of whom are school practitioners. These methods include narrative inquiry, phenomenology, ethnography, and case study research.

The boundaries among all types of qualitative research are more fluid and less distinct than those among different types of quantitative research. Qualitative research uses an **emergent design:** the methodologies may change throughout the study in order to better represent the reality of the persons and setting studied. A researcher might initially plan to conduct interviews but discover documents that provide rich details not covered in the interviews. However, all qualitative researchers begin with a preliminary plan that identifies the type of methodologies they expect to use. In addition, all qualitative research shares the following key characteristics:

- Studies are carried out in naturalistic settings.
- Researchers ask broad research questions designed to explore, interpret, or understand the social context.
- Participants are usually selected through nonrandom methods based on whether the individuals have information vital to the questions being asked.
- Data collection techniques involve methods such as observation and interviews that bring the researcher into close contact with the participants.
- The researcher is likely to take an interactive role through which she or he gets to know the participants and the social context in which they live.
- Hypotheses are formed after the researcher begins data collection and are modified throughout the study as new data are collected and analyzed.
- The study reports data in a narrative (using words rather than numbers) form.

Think about how David might carry out his study. How would his study illustrate these characteristics of qualitative research? The different types of qualitative research are varied. We first describe each type of qualitative research and then discuss the general steps for conducting qualitative studies.

Narrative Inquiry

The word *narrate* means "to tell a story in writing or in speech" (Friend & Guralnik, 1956, p. 975). **Narrative inquiry** is the study of the ways in which humans experience the world through examination of both the personal and social stories that they tell (Connelly & Clandinin, 1990). Narrative inquiry assumes that we all lead "storied lives" and that educators and learners are both "story tellers and characters in their own and others' stories" (Connelly & Clandinin, p. 2). In narrative inquiry, researchers describe the lives of people by collecting and retelling stories about their experiences and their interpretations of the meaning of these experiences. So if you enjoy reading and telling stories, narrative inquiry may be a research approach for you, although, as you might expect, the process of conducting a narrative inquiry study involves more than just storytelling.

Researchers interested in narrative inquiry may study a variety of phenomena. For example, research questions that might be examined through narrative inquiry include:

- What life experiences and turning points does an adult learner who begins an undergraduate degree after she is 40 years old describe as critical to her life changes?
- How do cancer survivors describe their experiences with family, friends, and medical professionals as they receive treatment?
- How does a first-year teacher in a small-town middle school describe the everyday events in her teaching?

Narrative inquiry has roots in many disciplines, including literature, history, psychology, anthropology, sociolinguistics, sociology, and education. Connelly and Clandinin (1990) have provided comprehensive discussions of how narrative inquiry has been applied to educational settings in a series of articles and books that are listed in the Suggested Readings at the end of this chapter. Some of the earliest uses of narrative inquiry in education included Goodson's historical account of teacher life histories (cited in Connelly & Clandinin, 1990), studies of adult and second language learning, classroom- or school-based narratives of teaching and learning (Clandinin & Connelly, 1988), and feminist studies of women's experiences (Personal Narratives Group, 1989). More recently, narrative inquiry has been used to study teachers' knowledge, reflections, and expertise and to "give voice" to their experiences (Creswell, 2002). In addition, narrative inquiry has been used to study the experiences of school psychologists (Tramonte, 2001), principals (Craig, 2003), counselors (Wong-Wylie, 2006), parents (Runswick-Cole, 2007), and students (Xu, Connelly, He, & Phillon, 2007) as well as educators in nonschool settings (Covin & Dirkx, 2003).

Types and Characteristics of Narrative Inquiry Many different types of studies fall under the heading of narrative inquiry. Table 6.1 presents descriptions of the most common types of narrative inquiry.

TABLE 6.1 Types of Narrative Inquiry

Type of Narrative Inquiry	Description
Autobiography	An account of a person's life written by that person
Autoethnography	A personal narrative written by a researcher who describes and interprets his or her own life story and experiences within a social, political, or cultural context
Biography	An account of a person's life written by another person (researcher)
Life stories or life histories	Narrative descriptions of the lives of one or a small number of individuals within a particular social or cultural context. It may be from birth to death or focus on specific important events.
Memoirs	An account of selected parts of a person's life written by that person; memoirs are autobiographical but are more selective and flexible than autobiographies
Narrative ethnography	Studies combining ethnography and life history by examining the lives of an individual or small number of individuals while the researcher engages in long-term involvement in a culture or community
Narrative performance	Dramatic reenactment of stories collected through qualitative research methods
Oral history	An oral record and interpretation of historical information based on the personal experiences of one or more persons
Testimonios Frequently used by Latin American activists in revolutionary movements	"An explicitly political narrative that describes and resists oppression of a particular group" (Chase, 2005, p. 653)

Sources: Chase (2005); Denzin and Lincoln (2005); Schwandt (2007).

Creswell (2002) points out that one can organize these subtypes of narrative inquiry based on three major questions:

1. Who writes the story?
2. Who or what provides the story?
3. How much of the person's life is recorded and presented?

All types of narrative inquiry emphasize strong relationships between researcher and participants; in fact, many narrative researchers use terms such as *listener* (instead of researcher) and *narrator* (instead of participant) to emphasize the deeply interpersonal nature of their interactions. Writing tasks in narrative inquiry are often shared by researchers and participants. However, some types of narrative inquiry are defined by who is the major author of the story. For example, an **autobiography** is the life story of a person as written by that person, while a **biography** is usually written by a researcher, often in consultation with the participant or focal person of the story. Sometimes the roles of researcher and participant are combined, as in **autoethnography**—personal narratives written by researchers who describe and interpret their own personal stories and experiences within a social, political, or cultural context.

The sources of information used in creating the narrative may be used to define different types of narrative inquiry. **Oral histories** are usually based on in-depth personal interviews, while biographies, personal narratives, and personal documents might also use diaries, journals, and letters as data sources. **Memoirs** and autobiographies focus on the participant's own experiences, while **life stories** and biographies might include interviews with other persons who know the focal person.

The duration of the story varies in different types of narrative inquiry. Life stories and life histories usually describe events from birth to the present, although they may focus on specific life-changing events that narrative researchers call *epiphanies*.

Narrative inquiry includes a diverse range of approaches, and there are some common key characteristics that define these approaches:

- *Narrative inquiry focuses on the experiences of individuals.* The focus of narrative inquiry is on understanding individual people and their stories. Social, cultural, and political context are included to provide deeper understanding of the stories of the individuals. However, these individual stories are considered to be important in their own right, whether or not they illuminate more general issues. Typically narrative inquiry studies focus on just one or a few individuals.
- *Narratives are organized so that they tell a story, usually in some type of chronology.* Narratives typically have a beginning, middle, and end. Some are organized in terms of past, present, and future.
- *Stories are collected, organized, and analyzed by researchers-listeners and participants-narrators working collaboratively.* Narratives are not simply first-person accounts of life events; they are organized and analyzed so that the meaning of the events to the narrator is made clear. This analysis occurs through deeply personal interactions between the researcher-listener and participant-narrator. In some types of narrative inquiry, the roles of researcher and participant are combined in one person.

- *Individual narratives are restoried to identify major themes and meanings.* Data from interviews, observations, and documents are analyzed and combined to form a coherent and meaningful narrative.

Theoretical Assumptions Underlying Narrative Inquiry Qualitative research is always shaped by the theoretical framework of the researcher. Chase (2005) highlights five theoretical assumptions that guide narrative inquiry. As a form of qualitative research, narrative inquiry emphasizes that narratives are not just stories that are true or false. Creating a narrative involves “retrospective meaning making—the shaping or ordering of past experiences” (Chase, 2005, p. 656). In other words, narratives are ways in which we make sense of our lives, selecting events that we perceive as important, organizing these events into meaningful wholes, and perceiving consequences of our actions over time.

Narrative inquiry also emphasizes that narrative can be a way to take action, an approach to research that falls under the advocacy or liberatory framework discussed in Chapter One. A narrator can shape his or her story to entertain others or to challenge the status quo. Narrative inquiry researchers attend to both what narrators say and how they say it in order to preserve the unique voice of each person and consider its social purpose.

Like all qualitative research, narrative inquiry emphasizes the social, political, and historical context within which a given narrative occurs. One’s story may be “enabled or constrained by social resources and circumstances” (Chase, 2005, p. 657). For example, narratives about women’s lives emerged at a time when the feminist movement supported the importance of giving voice to women. The context may help to better understand the lived experiences of a given person. In addition, individual stories may provide a more fine-grained understanding of how social, political, and historical forces develop and change. Narratives about individual slaves provided a deeper look into the early development of the civil rights movements.

Narratives are socially situated within the specific context in which they are told. Imagine how your own story of what your typical school day looks like might differ depending on whether you are talking to your best friend, a professor, or your mother! Researchers recognize that stories are flexible and shaped by the interactions between the narrator and audience or listener.

Finally, all qualitative research involves **subjectivity** or examination of how the researcher’s own experiences, biases, and assumptions influence the research. Throughout the process of data collection and analysis, narrative researchers examine the meanings and interpretations that they bring to narratives as they retell or restory them. Researchers also examine their own voices used in the writing process and consider how they shape the final narratives. This reflects the social constructivist assumption that reality is socially constructed with multiple interpretations possible. Table 6.2 summarizes the assumptions underlying narrative inquiry.

Phenomenological Research

Phenomenological research is the "study of everyday, lived experiences and the meanings that people construct from them" (McClelland, 1997, p. 108). Like narrative inquiry, phenomenological research looks closely at an individual's interpretation of his or her experiences. Phenomenologists also attempt to understand the meaning of experiences from the perspective of the participant. They recognize that there are many different ways to interpret the same experience and never assume that they (the researchers) know what things mean to the people they study. While the focus on how individuals interpret the meaning of their experiences is a characteristic shared by both narrative inquiry and phenomenological research, phenomenological researchers do not always aim to produce a chronological sequence of events that retells an individual's story. They also differ from narrative researchers in that they always seek out both individual aspects and shared commonalities of a focal experience that several groups of individuals have had. For example, a researcher might examine the stresses experienced by a group of single parents to see how the lived experiences of the parents are similar and different.

Research questions for phenomenological research identify a specific experience and seek both detailed descriptions of the experience as well as deep psychological explorations of the meaning of that experience to the person. Some sample research questions for a phenomenological study might include:

- "What is it like for parents to know and be known by others in rural schools" (McClelland, 1997, p. 108)?
- What are the shared and unique experiences of heterosexual persons who become allies/supporters of persons who are gay, lesbian, or bisexual (Kennedy & Voegtle, 2000)?

TABLE 6.2 Theoretical Assumptions Underlying Narrative Inquiry

1. Narratives are stories that involve retrospective meaning making as narrators reorganize and reinterpret past events.
2. Narratives are a means of taking actions or affecting listeners.
3. Social, political, and historical contexts influence how and what stories are told.
4. Narratives are socially situated, and the immediate social context affects how stories are told.
5. Narrative researchers must examine how their own experiences, biases, and assumptions influence the narrative and the voice used to tell it.

- How do urban teen mothers describe their current experiences and their expectations, goals, and future "possible selves" (Klaw, 2008)?
- What are the experiences of women who are sexual abuse survivors in dance therapy (Mills & Daniluk, 2002)?

Because phenomenologists appreciate that experience is varied and complex, they usually collect extensive amounts of data over time from their participants. Prior to data collection, phenomenologists reflect on their own experiences so that they can empathically enter into the world of their participants and become more aware of their own biases and assumptions. Data collection typically involves several group or one-to-one interviews with participants to allow them to reflect on their experiences and to add to what they said in earlier interviews. Whereas many qualitative researchers use multiple interviews, the phenomenologist focuses more on the essence of the human experiences and relies heavily on in-depth interviews as the most unbiased way to understand what the experiences mean to participants. According to Patton (1990), "The assumption of essence, like the ethnographer's assumption that culture exists and is important, becomes the defining characteristic of a purely phenomenological study" (p. 70). Summarizing our discussion so far, defining characteristics of phenomenological research include:

- Phenomenological research focuses on the essence of lived experiences of persons and the meanings that they see in these experiences.
- Phenomenological research uses multiple, in-depth interviews as its primary method of data collection.
- Phenomenological researchers use self-reflection to examine their subjectivities.
- Phenomenological researchers attempt to see the world through participants' eyes and empathically understand the meaning of their experiences.
- Phenomenological researchers try to identify both individual and common elements in the experiences of their participants.

Theoretical Assumptions for Phenomenological Research Phenomenological research methods grew out of the philosophical frameworks of existentialism and phenomenology, which emphasize that humans seek meaning from the experiences in their lives. Existentialism and phenomenology are frameworks that are sometimes described as *working from the inside to the outside*—that is, they examine deep psychological processes to illuminate external behaviors and actions. Contrast this with scientific realist approaches that always start with behavior and then infer people's intentions or motivations. Like many other qualitative approaches, phenomenological research would be classified as taking a social constructivist approach in that it assumes that there are multiple realities and that reality is socially constructed.

Although phenomenologists believe that people seek meaning in their experiences, they point out that most ordinary, everyday experiences are lived in a straightforward and unreflective manner. We do not always examine the subjective realities and meanings of our lives; on most days we simply live and experience life without much reflection. However, the meaning of our experiences is rooted in the details of these experiences and can be examined by a researcher who engages in meaningful dialogues with participants about these experiences. Therefore, the phenomenological researcher has two goals: first, to create a detailed description of the particulars of participants' experiences and, second, to encourage participants to examine the subjective meanings of their experiences.

Like narrative researchers, phenomenological researchers believe that experiences are situated in a particular social, historical, and cultural context that plays a role in how participants derive meaning from their lives. So it is important to describe the context and examine how participants understand it. The researcher's role in phenomenological research involves subjectivity that must be examined. Phenomenological researchers often engage in deep reflection about experiences of their own that are similar to those of their participants and examine their own preconceptions regarding these experiences. In addition, they strive to listen empathically to participants and seek to enter their lived worlds by soliciting as much detail about the experiences as possible. Table 6.3 summarizes the theoretical assumptions underlying phenomenological research.

TABLE 6.3 Theoretical Assumptions Underlying Phenomenological Research

1. Phenomenological research is based on existential and phenomenological philosophies and assumes that humans seek meaning in their lives.
2. Phenomenological research takes a social constructivist approach in assuming that there are multiple realities and these realities are socially constructed.
3. Phenomenological research emphasizes that social, political, and historical contexts influence people's experiences and the meanings that they derive.
4. One must examine the details of everyday life to get at deeper, subjective meanings.
5. Phenomenological researchers examine their own experiences, biases, and assumptions and attempt to listen empathically and enter the lived world of participants.

Ethnography

Ethnography is one of the oldest types of qualitative research. Like narrative inquiry and phenomenological research, ethnography has borrowed ideas and techniques from disciplines like sociology and anthropology and it uses inductive methods in collecting data.

The word *ethnography* is derived from the Greek words *ethos* ("tribe") and *graphos* ("something that is written"). Literally, then, ethnography is the science of writing about tribes or, to use more contemporary language, writing about cultural groups. Ethnographic researchers hope to provide rich narratives or descriptions of the communities or cultures under investigation (Miles & Huberman, 1994).

The overarching purpose of **ethnographic research** is to discover the patterns of a culture and its unique complexities in order to "paint a portrait" of the group, its interactions, and its setting. According to LeCompte and Schensul (1999), ethnography is a research method useful in the discovery of knowledge that is embedded within a culture or community. There is no consensus about exactly what *culture* means, but most sociologists and anthropologists believe that culture refers to attitudes, knowledge, values, and beliefs that influence the behavior of a particular group of people. Embedded in this definition is the recognition that shared experiences and interactions shape the individuals within the group. Ethnographers hope to explore these shared experiences by understanding what people do, say, and believe. This, of course, necessitates that the researcher develop a close relationship with the participants in the study. This relationship is enhanced by data collection procedures that bring the researcher and participants into close communication. Often, ethnographers (and case study researchers) rely on **key informants** or persons who are knowledgeable and who can provide the richest insights into the culture of the group and the issues addressed in the study. Ethnographers, like all qualitative researchers, typically talk with key informants many times throughout the study, checking on how the informants perceive and interpret events or activities. Key informants may help the researcher learn the unwritten rules guiding interactions and communication in a group and may give advice on how to approach particular situations.

Ethnographic research reports are usually lengthy, often taking the form of books. A major goal for ethnographers is to provide a "richly detailed description" (what anthropologists call a *thick description*) of the situation, capturing the full complexity of the nuances in interactions, cultural practices, and beliefs of the group. Good ethnographic studies should help readers feel that they are living the experiences of the groups studied and see the world through their eyes.

An example of an ethnographic researcher is Jonathan Kozol, who has written many books about the lives of inner-city youth and families. In one study, he wanted to portray the struggles and dreams of families living in poverty in Mott Haven, a community in the Bronx, New York. To accomplish this, he immersed himself in ethnographic research, living in the community of Mott Haven for several months at a time. Kozol's investigation brought him close to the children, families, and religious and educational leaders in the community. Through his work in the community, participants came to trust Kozol and to share with him their innermost feelings and perspectives on being poor and of color. Some of his key informants included the children and their parents, teachers, the director of an after-school program, and her assistant. By using primarily observations, interviews, and the language of his participants, he provided an insightful view of life in an urban community (Kozol, 1995, 2000).

Ethnography requires a tremendous amount of time and personal commitment on the part of the researcher. It is a research method chosen when questions or topics are embedded in cultural complexities and the researcher wants to understand cultural reality from the perspectives of the participants. The following are examples of research questions that might be explored through ethnographic research:

- What types of strategies are used by children to create and defend social roles and social order in a preschool?
- What is the out-of-school curriculum (those issues faced by students outside of the classroom) experienced by adolescent males of color growing up in an urban community?
- How do Latino children from immigrant families whose English proficiency is limited deal with the challenges presented to them in school?

Types and Characteristics of Ethnography Ethnographic research originated in the 19th and early 20th centuries as anthropologists studied primitive groups by living with them and immersing themselves in their cultures. However, these early anthropologists maintained their professional stance and examined the cultures "objectively" as they resisted the urge to "go native." Sociologists later borrowed many of these techniques to examine groups in the United States and other industrialized countries who were marginalized by people in power, such as Polish peasants, delinquent youth, and unemployed persons in urban settings. In the period from 1950 to 1980, educational anthropologists began to use ethnographic methods to study schools and classrooms. They looked at schools as institutions with their own social structures and cultures and described classrooms as tribes with rituals and customs. They examined cultural processes within schools, such as educational decision making regarding curriculum, placement of students, instructional methods, and learning. In the past 20 years, debates have raged about the role of the researcher in ethnographical research and the theoretical assumptions underlying this research. Although ethnographers always take the role of participant-observers, the degree of involvement and the relationship of the researcher and participant vary. Table 6.4 lists some of the types of ethnographies. As you can see, these overlap with types of narrative inquiry. Whether one classifies a study as narrative inquiry or ethnography depends on the extent to which one focuses on culture and identifying broad cultural patterns rather than telling a story.

As you can see from Table 6.4, ethnographers have created many variations in ethnographic designs. The three most common are realist ethnographies, ethnographic case studies, and critical ethnographies.

TABLE 6.4 Types of Ethnographic Studies

Autoethnography	Researcher reflects on his or her own experiences within a cultural context.
Confessional ethnography	Focuses on the researcher's fieldwork experiences.
Critical ethnography	Studies a group that is marginalized and examines issues of social power and inequality within this group; goal of critical ethnography is to increase awareness of how marginalized groups are dominated, repressed, or ignored and to advocate for fairness, equality, and justice
Ethnographic case study	Analyzes a single group, activity, event, or process with detailed attention to the cultural context
Ethnographic novels	Fictional work that draws from an ethnographic study of the cultural patterns of a group
Feminist ethnography	Studies girls or women and the cultural practices that result in oppression or lack of power
Microethnography	Focuses on a specific aspect of a cultural group and setting in detail
Postmodern ethnography	Examines and challenges societal problems that are related to modern emphasis on progress and inequalities that marginalize some groups
Realist ethnography	Provides an objective description of a culture and often uses both quantitative and qualitative data

Sources: Creswell (2002); Denzin & Lincoln (2005); Gay, Mills, and Airasian (2006); Van Maanem (1988).

Realist ethnography aims to provide a detailed and objective description of a culture. The researcher stays in the background and reports data in as unbiased a way as possible. Participants' perspectives are represented through quotes, but the researcher maintains an "outsider" perspective. For example, a researcher might observe the decision-making processes used by a teacher selection committee. The researcher, who is not a member of the committee, would report the processes and procedures used in making the decision. This type of ethnography often reports both quantitative and qualitative data and some might consider it a type of mixed-methods research. Clifford and Marcus (1986) have challenged the assumption underlying realist ethnography that researchers can be objective. Consequently, most realist ethnographers engage in reflexive examination of their own assumptions and biases and attempt to study and represent many different perspectives in their studies.

Ethnographic case studies focus on a single group, activity, or process and seek to identify "the shared patterns of behavior that it develops over time" (Gay, Mills, & Airasian, 2006, p. 485). Ethnographic case studies use mostly qualitative methods and involve more researcher participation than realist ethnographies. For example, Carger (1996) described the experiences of a Latino family whose children were English Language Learners (ELL) in a U.S. school over the period of one year. Carger was a tutor for the focal student and described the family's interactions with the school and teacher. One of the patterns that she observed was that the school did not always recognize or support the needs of its students who were ELL. Since case studies are quite varied in approach and include forms that are not ethnographic, we discuss this approach in a later section of this chapter.

Critical ethnographers seek to study issues of social power, inequality, and victimization through the use of mainly qualitative methods. The goal of this type of study is to increase awareness of how the marginalized groups are dominated, repressed, or ignored and to advocate for fairness, equality, and justice. A good example of critical ethnography is the work done by Kozol (1995, 2000) discussed earlier. An important aspect of Kozol's many books written about his experiences in Mott Haven involves an examination of the unfair treatment of its residents.

Although the types of ethnography differ considerably, they share some common characteristics:

- *Ethnographic research identifies and studies culture sharing groups.* Culture sharing groups include at least two persons who interact on a regular basis in a common setting.
- *Ethnographic research seeks to describe and interpret the shared patterns of behaviors, beliefs, ideas, and languages of a cultural group and individual variations within a culture.* Examples of shared patterns might include ideas about inclusion, procedures

for deciding on tracking of students, processes for evaluating teacher quality, or differences in language used by a teacher versus a school psychologist to describe students. Typically the shared patterns are types of knowledge or practices that are tacit—that is, persons are often unaware of how their own ideas guide their behaviors, and they cannot easily verbalize their knowledge or practices. As Polanyi (1967) said, "We can know more than we can tell" (p. 4); the job of the ethnographer is to examine and describe this tacitly held knowledge and behavior.

- *Ethnographers engage in extensive field research in which observations are a major data source.* All ethnographers use in-depth, firsthand observations of daily behavior over a lengthy period of time. Although the time period may vary, it is typical for ethnographers to collect data for a year or more.
- *Ethnographers collect multiple types of data.* While observations are a major source of data, ethnographers also analyze documents and artifacts in the site and conduct both formal and informal interviews with participants.
- *Ethnographers examine the broader historical, social, political, and cultural context within which a group functions.* Both realist and critical ethnographers emphasize that any group influences and is influenced by a broader context. Hence both types of researchers analyze and interpret their data by considering these broader frameworks, which are used as lenses through which all behavior is viewed.
- *Ethnographic researchers examine their own assumptions, biases, roles, and relationships in the study to understand how these influence their analysis and interpretation of results.* Critical ethnographers probe these issues in depth throughout the study; however, even realist ethnographers recognize the importance of subjectivity, the process whereby researchers examine their own assumptions and biases.

Theoretical Assumptions Underlying Ethnographic Research Early ethnographers might be described as having taken a scientific realist perspective because of their emphasis on objectivity and collection of both quantitative and qualitative data. However, a critique of realist ethnography by Clifford and Marcus (1986) raised issues that are still being debated. They emphasized that multiple perspectives exist in any social situation and the researcher's voice is only one of many that needed to be heard. They also questioned whether any study could be objective and said that the historical and cultural context for both the researcher and participants needs to be examined in terms of interacting theoretical frameworks and social forces such as feminism, racism, disability awareness, and societal power differentials.

Ethnographers today continue to debate many of these issues, and as qualitative researchers they stress the importance of making one's theoretical and philosophical assumptions clear. Most ethnographers use theoretical frameworks that draw from either the social constructivist or advocacy and liberatory approaches. Ethnographers taking a social constructivist approach emphasize the need for researchers to reflect on their own assumptions and biases and document these throughout their study. They also believe that researchers must build close and trusting relationships with participants so that they can fully understand the realities of the participants' lives. Since social constructivists believe that reality is socially constructed with multiple perspectives possible, they expect that the realities uncovered in their studies will be complex, multilayered, and sometimes conflicting.

Critical ethnographers are more likely to take an advocacy or liberatory approach. They believe that global changes such as increasingly diverse demographics in most countries, a worldwide economy, and international politics have contributed to systems of power and authority that affect all individuals and groups. They view any social situation in terms of issues such as power, empowerment, equality, dominance, repression, social struggle, and/or victimization. They view their role as an advocate for marginalized persons and groups with little power and, as you will see in our next chapter, this affects how they collect, analyze, and present their data as well as how they create, maintain, and withdraw from relationships with participants.

Case Study Research

Case study research is a form of qualitative research that endeavors to discover meaning, to investigate processes, and to gain insight into and in-depth understanding of an individual, group, or situation. According to Smith, as cited in Merriam (1998), case studies can be differentiated from other forms of qualitative research by the fact that these studies focus on a "single unit" or a **bounded system**. What, you might ask, constitutes a bounded system? According to Merriam, boundedness can be determined by asking "whether there is a limit to the number of people involved who could be interviewed or a finite amount of time [for observation]. If there is no end (actually or theoretically) to the number of people who could be interviewed or to observations that could be conducted, then the phenomenon is not bounded enough to be a case" (pp. 27–28). Stake (2005) points out that boundedness is defined in terms of the features that identify the case setting or group and in terms of activities that connect them. If a researcher conducts a case study of her inclusive preschool, the study's boundedness would be defined by the physical setting for the preschool, the persons who attend it or work with it, and the types of activities within it.

Although case studies are usually considered to be a type of qualitative research, they often include both quantitative and qualitative data. Stake (2005), a major case study researcher, admits that "case study research is neither new nor essentially qualitative. Case study is not a methodological choice, but a choice of what is to be studied" (p. 443). Essentially, the case defines the topic for the study, and any data that help to understand the case can be collected. However, case study researchers typically use many qualitative data sources.

One confusing aspect for students is that the word *case* is used frequently in the research literature, and it is often applied to studies that are not case studies. This is because the word *case* simply can be used to refer to a single group, individual, setting, or process. Hence ethnographers may refer to their site as a case. Single subject researchers also sometimes refer to their studies as single case studies.

Case studies are used in many different fields by both professional researchers and practitioners including historians, medical practitioners, program evaluators, professional educational researchers, educational practitioners working in teams in school settings, and institutional researchers studying their own organizations. Research questions can probe practical or broad research issues, although the focus is always on what one can learn from the specific cases studied rather than making theoretical generalizations. A case study could be proposed if you are conducting a study that gets you close to a particular individual, group, school, classroom, program, or event. As in ethnographic research, your goal would be to provide a richly detailed description (a thick description) of the situation, to capture the full complexity and uniqueness of the case information. Examples of research questions for case studies include:

- How do parents, teachers, and students in a new inclusive preschool learn to work together?
- How does a middle school music teacher incorporate students' cultural backgrounds into her curriculum and instruction?
- How do the roles of school psychologists, administrators, and special education teachers change when a school uses a data-based, problem-solving model to make decisions about placement in special education?
- What types of resources or supports facilitate persistence in high school students with fetal alcohol syndrome?
- How does a new school administration program prepare six school principals to be leaders in the 21st century?

Types and Characteristics of Case Studies Case studies vary enormously in length, complexity, focus, and purpose. Some focus on a single individual and may last only a couple of months. Others involve multiple sites that are compared for deeper understanding of a broader issue. However, most case studies can be classified as being intrinsic, instrumental, or collective case studies.

Intrinsic case studies seek understanding of a specific case (for example, person, program, school, or activity) that is considered to be important in its own right. The researcher may select the case because s/he has a preexisting relationship with the student, program, or school, or the researcher may select a new case because of its atypical and unusual nature. Intrinsic case studies are often more descriptive than interpretive, focusing on the specific characteristics of the person or program studied, the context in which the case is embedded, and the unique processes involved in the case. The goal is to understand the case in more depth so that the researcher can learn about it in depth. A study of an inclusive preschool might be an intrinsic case study if the researcher's goal is to learn about this particular school because she hopes to work there. Intrinsic case studies often offer a way to study individuals who are exceptions to the rule. For example, a large-scale quantitative study might report that overall a particular treatment is effective for a group of students, but there is variability in how well it works for some students. A case study could be conducted to see whether modifications of a treatment might make it more effective for a particular student or whether a different treatment is needed.

Instrumental case studies examine specific cases in order to gain insight into some broader issue. The case is studied in depth but the goal is to understand how the particulars of the persons or settings studied help one understand some broader issue on a deeper level. For example, Lau, Sieler, and Muyskens (2006) studied how one urban school district used a data-based, problem-solving model to make decisions about placement in special education. The details from this one district were used to discuss the broader issue of how decisions about student placement are made and whether data-based systems result in high student success. Instrumental case studies often challenge broad claims about teaching and learning by providing detailed examples of cases that do not fit with accepted generalizations. For example, Anderson (2008) reported detailed case studies of elementary children writing persuasively despite widely accepted claims that these children were capable of making persuasive argument orally but not in writing.

Collective case studies (also called *comparative case studies*) study and compare multiple cases in order to provide insight into an issue. The goal of a collective case study is to understand both the uniquely individual aspects of the case as well as their commonalities. For example, Duquette, Stodel, and Fullarton (2006) examined eight middle school students with fetal alcohol syndrome to learn what factors helped them to persist in high school. These researchers' case study reports both differences between the students and common factors such as the importance of supportive social networks that fostered persistence. Collective case studies are often used to identify characteristics of successful programs. Humphrey, Wechsler, and Hough (2008) collected extensive qualitative and quantitative data on alternative teacher certification programs at seven colleges to identify characteristics of the programs that contributed to their effectiveness.

As a research approach that draws heavily on qualitative data, case studies share the characteristics noted at the beginning of this chapter that are common to all qualitative research, such as naturalistic setting, interactive relationships between researcher and participants, purposeful selection of participants. Some characteristics that define a qualitative study as a case study follow.

- *Case studies focus on in-depth examination of one or multiple bounded cases.* Sounds circular, huh? The point is that case studies examine persons, programs, or processes in order to gain a deeper understanding of those specific persons, programs, or processes; the researcher may not be concerned whether there are any broader generalizations that can be made about the case. In fact, Stake (2005) cautions that if a case study researcher is too concerned with generalization, she or he may miss particulars of a case and fail to really understand it.
- *Case study research seeks to situate a case in its historical, social, and cultural context.* Like other qualitative researchers, case study researchers view this context as essential to understanding the individual case. For case study researchers, this includes the individual's history and social relationships prior to the start of the case study.
- *Details on physical setting and major participants in the case are presented.* Like ethnographers, case study researchers seek to create a realistic portrayal of a case setting. Therefore it is essential to describe both the physical setting as well as all persons who interact with the person, group, or program of interest.
- *Case study researchers collect multiple types of data, sometimes including extensive quantitative and qualitative data.* Quantitative data are often included for decision-making purposes in an intrinsic case study or to summarize cross-case comparisons in a collective case study. However, in most case studies, qualitative data are collected to probe the case in depth.

Theoretical Assumptions Underlying Case Studies The extensive use of qualitative data and analysis by case study researchers suggests that their theoretical assumptions might best be classified as social constructivist. Their emphasis on multiple perspectives, social and political context and extensive use of triangulation to ensure that a case study accurately depicts the reality of the participants are consistent with a social constructivist approach. Finally, case study researchers recognize the problem of researcher bias and strongly promote extensive researcher reflection as a control for bias, again supporting a social constructivist approach.

However, the enormous variability in case studies and the flexible use of both quantitative and qualitative data suggest that the theoretical framework most relevant for some case study research is *pragmatism*. Case study researchers certainly use any data source and research methodology that will help them to better understand their case. Intrinsic case studies also often seek to improve or at least better understand a specific educational situation or problem. At the very least, one can say that case study researchers are more pragmatic than any other type of qualitative researcher.

STEPS IN DESIGNING QUALITATIVE RESEARCH

As noted earlier, qualitative research uses an emergent design that may change throughout the study. Therefore, it is not possible or appropriate to specify a set sequence of steps in designing a qualitative study. The "steps" that are presented here may be repeated or reordered as you carry out a study. We discuss issues that are common to all qualitative research.

Step 1: Identify a Research Topic or Focus

In qualitative research, topics are typically identified by the researcher based on experience, observation in the research settings, and readings on the topic. Although topics are set at the beginning of the study, the focus of the study may change during the data collection phase. The topic may focus on a person, setting, or experience that will be probed in depth. For researchers taking an advocacy-liberatory framework, the topic is defined in relation to broad social, political, historical, or cultural issues. For example, Wohlwend (2009) looked at how the media-created "princess identity" depicted in films and toys entered into the identities created by girls playing in a kindergarten class. Her initial topic thus included gender and media as two broad social forces as a focus for her study.

Step 2: Conduct a Review of Literature

In qualitative research, the researcher reviews the literature to identify information relevant to the study, establish a theoretical framework, and write a research question. This literature review often continues while data are being collected and allows the researcher to refine the research question. There is some disagreement among qualitative researchers about the extent of literature review that should be done before the start of a research study. Those who argue for a limited review typically fear that if too much literature is reviewed, the researcher will enter the setting with preconceived notions about what she or he is investigating. Although it is important to have some knowledge about the issue being investigated, too much knowledge may reduce the researcher's openness to all possible realities. One goal of the literature review may be to identify limits to what is known about a particular topic (Wertz, 2005). For example, in their study of the experiences of victims of child sexual abuse who participated in dance therapy, Mills and Daniluk (2002) first described research on the problems that victims of child sexual abuse report and research on dance therapy. However, they point out the lack of research combining these two topics.

Sometimes the review of literature is conducted after data have been collected and analyzed so that the stories of the individual are considered first and then links to past research or literature examined. The review might also include literature on theoretical frameworks that can guide the study.

Step 3: Define the Role of Researcher

In all qualitative research, the researcher must decide to what degree she or he will become involved with the participants. In general, because of the nature of qualitative research, the researcher has close contact with the participants. Qualitative researchers seek to create respectful and close relationships with participants that involve either active participation in the participants' day-to-day activities or in-depth learning about their lives through observations and interviews. Researchers taking an advocacy-liberatory approach also seek to involve participants in decisions about data selection, organization of data, and writing the research report. Qualitative researchers play an interpretive role in the data analysis and writing of the report and by describing the social, cultural, and historical context for the study.

Step 4: Manage Entry Into the Field and Maintain Good Field Relations

Once the researcher has clearly defined the research topic or focus, a field of study (for example, a place to conduct the research) must be identified and contacts made to secure permission for the study. Due to the intensive and long-term requirements of data collection, convenience and access may be considerations, as well as how much can be learned from the site. The selected field must be consistent with the research topic. If the topic or focus is charter schools, the researcher must decide what constitutes a charter school and which charter school would best help to answer the research questions. Once researchers have identified the field of study, they must then prepare to introduce themselves and the nature of the study to the persons who control access to the site—often school administrators. This must be done with care and consideration, because the administrators likely will want to know exactly what the study is designed to do and how both the study and the researcher's presence will influence the day-to-day functioning of the school. It may also be necessary to obtain permission from several different persons. For example, as part of his ethnographic doctoral thesis, one of our colleagues enrolled as a student in a high school mathematics class because he felt that sharing classroom experiences was the best way to get to know the students and to understand what they wanted to get out of school. To enroll in the class, he had to obtain permission from the instructor and the principal of the school. Once institutional permission is granted, you must obtain all the necessary permissions from those who will be participating in the study (and from guardians or parents if the participants are younger than 18 years). If one is studying multiple sites, as in a collective case study, permissions must be obtained for each site and gatekeepers identified.

In addition to obtaining permission, researchers must also identify **gatekeepers**— persons who officially or informally facilitate access to persons or places. Gatekeepers may suggest persons for interviews or convey inside information about times and places for observations that will provide deeper insight into a setting.

Sometimes, access may be facilitated because researchers have a preexisting relationship with individuals in the setting. However it is still important to be clear about the types of research activities that will be conducted and the time frame for the study. Maintaining trusting relationships throughout the study can sometimes be complicated by the researcher's relationship with the participants. In some studies, counselors, teachers, administrators, or school psychologists might be involved both as practitioners working with students and as members of a research team. It is essential that everyone is clear about what data are being collected for research, how privacy and confidentiality are respected, and how the data will be used. At no time should the research interfere with regular instruction or services; it is hoped that the research will serve to enhance instruction or services.

After the researcher gains entry into the field, he or she must establish and maintain good field relations. Researchers must consider how to build trust and rapport with participants over time and establish credibility so that participants will act naturally with them and be willing to share their experiences. It may take time before conversations and interactions are "natural" and participants have become adapted to the researcher's presence.

The establishment and maintenance of good field relations requires that trust and credibility be established. Sensitivity, honest communication and nonjudgmental interactions are critical characteristics of a good researcher and a necessary part of good field relations. Length of time in the field also promotes strong field relationships, since this connotes concern about deep understanding of a setting to participants. Qualitative research often includes data on deeply personal and sensitive issues, and attention to confidentiality is essential to good field relationships.

Step 5: Write Qualitative Subquestions

Qualitative subquestions are designed by the researcher and are based on the topics or research questions identified both at the start of the study and as the study progresses. Qualitative subquestions help the researcher to focus data collection and allow the data collection to proceed in a systematic way, but they should not predetermine what the researcher will find. For example, David from this chapter's opening vignette might have written the following qualitative subquestions:

- What are the experiences of students who are ELL and their families when the students first enter a large urban middle school?
- How do these experiences change or not change over time?
- What events or factors do students and families identify as important in their experiences, and how do they describe these?

As you can see, these questions are designed to guide the researcher as she or he proceeds with data collection procedures, but they avoid making assumptions about the types of experiences that will emerge.

Step 6: Select Participants

As noted in chapter 5, participants for qualitative research are selected through purposeful sampling. The researcher examines his or her qualitative subquestions and selects a purposeful sampling strategy to select participants who are best able to provide the information essential for the study. For example, David might select two families with whom he has a close relationship and who have experienced the most difficulties with the school. He might also want to select families who have been in the country for at least a year so that he can examine changes in their experiences. However, the most important consideration in sampling for any qualitative study is that the individuals have information or experiences related to the research questions that they are willing to share. (See Table 5.3 in chapter 5 for a summary of purposeful sampling strategies.) Qualitative researchers typically select only one or a small number of individuals to study. Because researchers emphasize that each individual's experiences are unique, they are most likely to focus on strategies such as critical case or extreme case—the approach described for David.

Selection of participants often occurs in multiple stages and may occur spontaneously. For example, in ethnographies and case studies, one must select a setting to study and then select participants, documents, and artifacts must be. Typically, the researcher has an initial plan for selecting who or what to study. Purposeful selection strategies such as *maximum variation sampling* may be used to obtain a variety of perspectives. Persons could be selected because they differ in gender, age, family structure, location, or other variables. Remember, however, that the goal is *not* to select a representative sample; samples for qualitative research are almost always too small to generalize findings. The researcher seeks variation in participant perspectives, recognizing that the small sample makes it impossible to know how widespread the patterns observed are to persons outside the study. After the researcher enters the field, additional selection decisions may need to be made. *Opportunistic sampling* may be used to select persons or documents encountered at later points in time if these are seen as conveying useful information. *Snowball sampling* may also be used if initially chosen participants can nominate others who can provide information relevant to the research questions.

How does a qualitative researcher decide when enough persons and data sources have been collected? There is no table of minimum size samples for qualitative research. Qualitative researchers often continue to seek out participants until they have reached **saturation—** a point at which the researcher subjectively decides that adding more participants will not provide new insights into the topic. Typically this decision is made as the researcher analyzes data and determines that new data are very similar to that already reported. It is often challenging to find participants for qualitative research, because there may be no list of persons who have had a particular experience, such as the victims of sexual abuse who have participated in dance therapy in Mills and Daniluk's (2002) study. The researchers used newspaper advertisements and worked with dance therapists to identify potential participants. In McClelland's (1997) study of parents in rural schools knowing and being known by others, he used the snowball strategy in which parents who were interviewed first nominated others for the study.

Step 7: Collect the Data

Next, the researcher moves on to collecting data. Data collection in qualitative research generally includes observations, interviews, and document analyses. Protocols are created to provide broad guidelines for data collection. Protocols may be revised or new ones created as the study progresses. Qualitative researchers typically include more than one data collection technique or use the same technique at different times to validate findings. These different data sources are later compared with one another in a process called **triangulation**. For researchers taking an advocacy-liberatory approach, data collection and analysis involve close collaboration with participants to ensure that the data collected accurately depict their experiences. As data collection nears completion, qualitative researchers must also consider how to withdraw from the study site. Researchers should let participants know that they will be leaving the site and provide ways to continue communications. Due to the intensive commitments of time and energy, many qualitative researchers look for ways to reciprocate and show their appreciation to participants.

Step 8: Analyze and Interpret the Data

In all qualitative research, data analysis and interpretation are continuous throughout the study, so that insights gained in initial data analysis can guide future data collection. Data in qualitative research are analyzed through the reading and review of data (observation notes, interview transcripts) to detect themes and patterns that emerge. The researcher summarizes and explains the results by describing the major ideas, patterns, or themes that emerge from the analysis. Interpretation may also involve discussion of how the findings from this study relate to findings from past studies in this area. The researcher summarizes his or her own interpretation of the data and provides thick descriptions from which the readers can draw their own interpretations. More information on qualitative data analysis and interpretation is provided in chapter 7.

Step 9: Disseminate Results

Qualitative researchers share their findings with other professionals through journals, reports, Web sites, and presentations at formal and informal meetings. Our students note that although qualitative studies are fun to read, they are frequently quite long! Due to the voluminous amount of data collected, reports on qualitative studies are sometimes published as books. Qualitative researchers frequently use creative formats for reporting results, inviting their audiences to enter into the experiences of participants by witnessing dramatic performance or artistic representations. Qualitative researchers taking an advocacy-liberatory approach often ponder how their results can be written and disseminated in a way that promotes social change. For example, David will need to think about how he can share the results from his study in a way that will change the harmful attitudes and practices that he sees in his school.

Table 6.5 summarizes the steps in conducting qualitative research and discusses how the steps differ for different types of qualitative research.

TABLE 6.5 Steps in Conducting Qualitative Research

Step	Narrative Inquiry	Phenomenological Research	Ethnography	Case Study
1. Identify a research topic or focus	Research topic describes the individuals whose stories will be collected.	Research topic describes the experiences to be examined.	Research topic describes a group, process, or setting that will be examined.	Research topic describes one or more groups or settings and the issues to be examined.
2. Conduct a review of literature	Literature review is deemphasized. Review might focus on prior research on persons who have lived through similar experiences or research using methodologies useful in the proposed study.	Literature review deemphasized. Review might include sources to prepare researcher for observations and interviews as well as researcher's reflections on his or her own experiences related to the topic.	Literature review includes studies of similar settings or groups and theoretical frameworks for understanding cultural, social, and political issues.	Literature review focuses on studies of similar cases, information pertinent to the case and examples of other case studies.
3. Define the role of the researcher	Researcher tries to reduce the distance between selves and participants. Collaboration between researcher and narrator is essential at all phases of the research.	Researcher is primarily a listener who prepares interview questions and facilitates interactions.	Researcher tries to develop an "insider" point of view by participating actively and frequently in the day-to-day life of the group while maintaining awareness of how his or her own history and culture shape experiences within the study.	Close and long-term relationships are established with participants, especially in intrinsic case studies in which researcher may have a previous relationship with participants.
4. Manage entry into the field and maintain good field relations	Researcher works to develop close emotional ties to a person or group with a shared set of experiences.	Researcher works to develop close emotional ties to a group with a shared set of experiences.	Researcher gradually enters field and spends extensive amounts of time becoming part of the community.	Researcher works closely with a well-defined group in order to understand people, activities or programs within the community.

5. Write qualitative subquestions	Researcher focuses on how individuals perceive the meaning of their experiences and how they relate these meanings through stories.	Researcher focuses on commonalities and individual differences in how individuals describe the meaning of their experiences.	Researcher identifies the group, setting, issues, and processes to be examined in the study.	Researcher identifies the types of interactions, materials, behaviors, or issues to be examined in the cases.
6. Select participants	Typically one or a small number of participants are selected. Use of typical, critical, and extreme case purposeful sampling is common.	Typically one or a small number of participants who have a similar experience are selected. Use of snowball purposeful sampling or recruitment through groups, professionals, or ads is common.	Multistage selection process and opportunistic sampling is common. Persons, settings, documents and/or artifacts sampled.	Multistage selection process involving persons, settings, documents, and/or artifacts is conducted. Typical sample, intensity sampling, or maximum variation sampling is often used depending on goals of study.
7. Collect the data	Multiple interviews and conversations are major data sources. Other sources, such as journals, letters, or diaries, are collected to help tell the person's story.	Initial observations and deep researcher reflections may be conducted to prepare for interviews. Intensive multiple interviews over time are the major data source to represent the participants' realities.	Multiple types of data are collected, including observations, interviews, documents, and artifacts. Initial observations are broad, and later ones are more focused on emerging issues and patterns.	Multiple types of data are collected, including observations, interviews, documents, and artifacts. This allows deeper insight into case and triangulation. Sometimes quantitative data are also collected.

(*Continued*)

TABLE 6.5 (*Continued*)

Step	Narrative Inquiry	Phenomenological Research	Ethnography	Case Study
8. Analyze and interpret the data	Researcher reviews data multiple times to select and organize information for retelling the individual's story and establishing a chronology of events and ideas. Data analysis includes search for themes within each story or across participants and examination of strategies and voices used by participants to tell their stories.	Researcher reviews data multiple times to examine both the differences and the similarities in the experiences of participants and to probe more deeply into the essence of their experiences.	Data analysis occurs throughout data collection in order to guide future observations and identify areas for which additional data are needed.	Data analysis is continuous throughout the study. Researcher engages in "criss-crossed reflection," examining the case within different subsettings and from different perspectives. Often both quantitative and qualitative data are used to summarize results. Rich detail provided on the context, people, interactions, activities, and processes involved in the case; both unique and common elements in the cases are discussed.
	Reports are often written with participants and typically contain lengthy excerpts from interviews or other data sources.	Results may be disseminated through artistic performances or data summarized and compared to works of literature or art in an attempt to derive the deep psychological meanings embedded in experiences.	Results are often summarized in multiple publications or books because of extensive data collected.	Results may be presented through regular professional forums (for example, journals, Web sites, conferences) or to more limited audiences (such as faculty meetings or meetings between client and program evaluator).

EVALUATING NARRATIVE INQUIRY AND PHENOMENOLOGICAL RESEARCH

Although qualitative research studies make for good reading, students often are uncertain about how to evaluate whether the studies are "good research." Because many qualitative studies focus on in-depth description instead of providing evidence that certain outcomes have been achieved, students often ask, "What's the point of this study?" One common error is using criteria that are appropriate for quantitative research to judge the goodness of qualitative studies. For example, David's study of families whose children are ELL in one urban school might be criticized because his sample is not representative and the results could not be generalized to other schools. However, this argument ignores the major goal of qualitative research, which is to provide an in-depth understanding of a limited setting, group, or person. This in-depth understanding may be cultivated through detailed descriptions, but it also involves deep analysis of how educational processes occur and how certain outcomes are achieved. Criteria for evaluating qualitative research differ from those used in quantitative research in that they focus on how well the researchers have provided evidence that their descriptions and analysis represent the reality of the situations and persons studied. Many different criteria have been proposed for evaluating qualitative studies; and Lincoln and Guba (1985) provide extensive discussion of these criteria. Here we focus on four major areas and the types of evidence—credibility, dependability, transferability, promoting action and collaboration—that researchers might present to support the quality of their overall study and their analysis and interpretation of their data.

Credibility

Credibility refers to whether the participants' perceptions of the setting or events match up with the researcher's portrayal of them in the research report. In other words, has the researcher accurately represented what the participants think, feel, and do and the processes that influence their thoughts, feelings, and actions? Credibility parallels the criteria of validity (including both validity of measures and internal validity) in quantitative research, although qualitative researchers do not discuss extraneous variables in the assessment of credibility. Rather, they look at whether the researcher's methods are likely to yield accurate and deep pictures of the experiences of the participants.

Evidence of credibility can take several different forms. First, a good qualitative study discusses how the researcher engaged in repeated, prolonged, and substantial involvement in the field. The amount of time devoted to data collection will depend on the nature of the study, but all qualitative studies should indicate how much time was spent in the field and how the researcher established and nurtured strong relationships with the participants. In general, one would expect a researcher to spend more time in the field in an ethnographic study than in a case study. In both types of study, taking part in meaningful interactions with participants enhances credibility. Since both narrative and phenomenological research seek the deep meanings of experience—not just surface details—participants may find it difficult to become aware of or describe these subjective aspects of their experiences. Credibility might be examined by asking the following questions:

- Did the researcher conduct multiple interviews with the participants, and did these interviews probe their experiences in depth?
- Is it likely that the participants were honest in describing their feelings and experiences?
- Did the procedures used by the researcher promote deep reflection and analysis of the meaning of the experiences?
- Is the story told by the participants authentic?
- Is the description too selective, omitting difficult or embarrassing incidents and ending with an overly optimistic tone?

A second aspect of credibility involves checking on whether the researcher's interpretations of the data are valid. Narrative researchers emphasize that the researcher shapes the final story that is told; it is critical to ensure that the retelling of the story by the researcher captures the true meaning of the participants' experiences. Chase (2005) cites a study in which the counselors of women who experienced spousal abuse were observed to transform the women's stories into formulas that fit the counselors' theories of domestic violence. The women in counseling did not see themselves as denying their victimization, although their counselors transformed their stories into tales of denial. The credibility of this study was compromised by the inability of the counselors to capture the true reality as described by the women in their study. Similarly, phenomenological researchers interpret the participants' descriptions of the meaning of their experiences. Hence narrative and phenomenological researchers both emphasize that researchers need to engage in analysis of how their own thoughts and stories change in the process of conducting research. In ethnographic and case study research, summaries of observations and document analysis are interpreted by researchers whose biases might color what they see. Therefore, continuous examination by the researchers of their own thoughts and feelings throughout the study is critical.

All qualitative researchers collect multiple sources of data to ensure that they have a deep understanding of the phenomena being studied. To enhance credibility, the researcher discusses how the information provided by these different data sources was compared through triangulation to corroborate the conclusions. If the sources provide conflicting information, the researcher should discuss possible reasons for the conflict. Another way to provide evidence supporting the hypotheses that emerge during qualitative research is **negative case analysis**. Negative case analysis involves examining the data for examples that contradict or disconfirm the hypothesis. When negative instances are identified, the researcher should revise the hypothesis or provide an explanation of why the case does not fit. In narrative research, this does not mean that one looks for conflicts in the stories told by participants. Rather one examines whether different data sources support the researcher's interpretations of the data and the retelling of the participants' stories. Similarly, in phenomenological research, a researcher might examine the interviews of each participant to see if there are any instances that negate something identified as common to the experiences of all participants. In ethnographic or case study research, the researcher's interpretation of cultural rules and customs can be examined by looking for instances when these rules or customs do not seem to apply. If such negative instances are found, the researcher must revise the description of either the shared aspects of the participants' experiences or the cultural rules or customs.

Because qualitative researchers often take a social constructivist or advocacy-liberatory framework, they may not expect that all participants will share the same perspectives. So it is essential to seek out and present a balanced view of all possible perspectives. To ensure that the researcher's own biases do not influence how the perspectives are portrayed, many researchers use **member checks,** in which the transcribed interviews or summaries of the researcher's conclusions are sent to participants for review. In addition, researchers continually monitor their own subjective perspectives and biases by recording reflective field notes or keeping a journal of their thoughts. Another strategy is to arrange for a **peer debriefer**, a colleague who examines the field notes and meets with the researcher on a regular basis, asking questions to help him or her reexamine assumptions and consider alternative ways of looking at the data. In narrative research studies, continuous collaboration between researchers and participants may serve the function of peer debriefing, since participants are often viewed as equal partners in research.

Qualitative researchers always assume that individual events and experiences are situated in social, historical, and cultural contexts that shape these events and experiences. Therefore, a full and accurate interpretation of a study must include a detailed description of social, historical, and cultural contexts relevant to the research study. This context may be provided through review of literature or information obtained in interviews or documents. The question is, Does the study consider how the experiences of the individuals shape and are shaped by their contexts?

Researchers taking an advocacy-liberatory framework may also insist that evaluation of a study include **attention to voice** (Lincoln, 1995). The questions are

- Who is speaking for whom?
- Does the researcher seek out and engage persons who are least empowered in the society?
- Are these voices given detailed coverage and a prominent place in the analysis and final report?

Narrative researchers often ask, "Who owns the story?" They may even make participants primary or secondary authors of the study. At the very least, they examine the final research report to see whether the writing reflects the voice of the researcher or participant or both. Ethnographic and case study researchers show strong attention to voice by using extensive quotes and including the participants' own language and style of speaking as part of the report writing. Studies taking an advocacy or liberatory approach may also insist that data be examined from a theoretical framework such as feminist theory.

A final and demanding strategy used by some qualitative researchers to support credibility is an **external audit**. In an external audit, an independent researcher examines all of the data collected in a study with the following questions in mind:

- Are the findings grounded in data? Is there a clear connection between each finding and some part of the data?
- Are the themes appropriate to the data? Are all interpretations and conclusions supported by the data?
- Have researcher biases been well controlled?

Dependability

Dependability is a criterion for qualitative research that parallels reliability in quantitative research, although it is not assessed through statistical procedures. *Dependability* refers to whether one can track the procedures and processes used to collect and interpret the data. Good qualitative studies provide detailed explanations of how the data are collected and analyzed. Recording devices such as audiotapes and videotapes are used extensively in all types of qualitative research to support dependability. Dependability is increased when research studies discuss how the relationship between the researcher and participants was nurtured and how the interviews or observations were structured. Dependability is often the difference between an experiential report that simply summarizes a researcher's conclusions and an empirical, research-based qualitative study that includes a thorough explanation of methods. Although it is not possible for qualitative researchers to include all of their data in their results, many qualitative researchers make their data available for review by other researchers.

Transferability

Although qualitative researchers do not expect their findings to be generalizable to all other settings, it is likely that the lessons learned in one setting might be useful to others. However, qualitative researchers insist that *readers* (not the researcher) must make this judgment, because the appropriateness of the lessons depends on the contextual similarities and differences between the site studied and the site to which the reader hopes to transfer its findings. **Transferability** refers to the degree of similarity between the research site and other sites *as judged by the reader.* Transferability is assessed by looking at the richness of the descriptions included in the study as well as the amount of detail provided about the context within which the study occurred. Because the reader is the person who must judge transferability, richly detailed or thick descriptions enable the reader to make judgments about the similarity of the participants, schools, resources, policies, culture, and other characteristics of the research site and the reader's own site. Transferability is *not* whether the study includes a representative sample; it is how well the study has made it possible for readers to decide whether similar processes will be at work in their own communities by understanding in depth how they occur at the research site. Therefore one would not ask if all allies of gay, lesbian, or bisexual persons have the same experiences as the allies in Kennedy and Voegtle's (2000) study. A more appropriate question would be whether the descriptions of these allies are detailed enough to allow readers to judge whether persons they know in similar settings might have the same experiences. Note that it is not essential for a qualitative study to be transferable. A person or setting may be studied because he or she or it is unique or illustrates an exception to a general pattern observed in other studies. The rich detail helps to illustrate how the case is different and allows future transferability if similar situations are studied.

Catalytic Authenticity

Action researchers and qualitative researchers taking an advocacy or liberatory framework often evaluate research based on whether it has stimulated action that will improve education or enhance the lives of persons with little power. Lincoln and Guba (1985) refer to this as **catalytic authenticity**, asking whether the research has stimulated change for the better in ways that are truly desired by the study participants. Evidence that this has occurred includes descriptions of how participants collaborated with researchers in determining what changes needed to occur and in planning any actions. Qualitative researchers often argue that the research process is likely to result in increased self-understanding in participants. However, some argue that the researcher must play an active role in ensuring that the stories and experiences described in the research are told to persons in power so that real social change occurs. Many researchers emphasize that the benefits and privileges of doing research should be shared with participants. This may take the form of sharing royalties from the publication of a book or obtaining funding to make it possible for research participants to take part in presentations at conferences. Educational conferences are increasingly enriched by the participation of teachers, students, and community members in presentations; many conference presentations use either narrative or phenomenological research approaches. If the ultimate goal of educational research is to improve education, the criterion of promoting action and collaboration is one way to evaluate whether the research was worth doing! Table 6.6 summarizes the criteria for evaluating qualitative studies and the methods for meeting these criteria.

TABLE 6.6 Criteria for Evaluating Qualitative Research Studies

Criteria	Methods to Meet Criteria
Credibility and control of researcher bias	Prolonged and meaningful participation in setting
	Triangulation of multiple data sources
	Negative case analysis
	Participant review of interview transcripts
	Member checks
	Peer debriefer
	Attention to voice
	External audit
Dependability	Detailed description of data collection and analysis procedures
	Use of videotape and audiotape
	Data made available for review
Transferability	Rich descriptions of setting, participants, interactions, culture, policies, and so on
	Detailed information on context and background
Promoting action and collaboration (catalytic authenticity)	Description of collaboration with participants
	Description of ways in which research changed the lives of participants
	Coauthorship of publications
	Sharing royalties and other benefits of publication

SUMMARY

Qualitative research is a methodology that has been borrowed from disciplines like sociology and anthropology and adapted to educational settings. Qualitative researchers use inductive methods of reasoning and believe that there are multiple perspectives to be uncovered. Qualitative researchers focus on the study of social phenomena and give voice to the feelings and perceptions of the participants under study. Qualitative studies are carried out in naturalistic settings, where researchers ask broad research questions designed to explore, interpret, or understand the social context and where participants are selected through nonrandom methods based on whether the individuals have information vital to the questions being asked. As part of the process, the qualitative researcher is likely to take an interactive role in which she or he gets to know the participants and the social context in which they live. Triangulation is a process used by qualitative researchers for data analysis when different data sources are compared with one another. In qualitative research, results from the study are reported in a narrative form. There are several types of qualitative research. Narrative inquiry research is the study of the ways in which humans experience the world through examination of both the personal and social stories that they tell. Researcher and participants work closely together to select, organize, and restory the experiences of participants in a narrative inquiry study. Phenomenological research looks closely at individuals' interpretation of their experience and attempts to understand the meaning of an experience from the perspective of the participants. Multiple one-to-one interviews with an individual are typically used with phenomenological research. Ethnography is a type of qualitative research in which the researcher tries to provide a rich narrative or description of the communities and cultures under investigation. The overarching purpose of ethnographic research is to discover the essence of a culture and its unique complexities and paint a portrait of the group, its interactions, and its setting. Case study research is a common form of qualitative research that endeavors to discover meaning, to investigate process, and to gain insight into and in-depth understanding of individuals, groups, or situations. Case studies are differentiated from other forms of qualitative research by the fact that they employ a bounded system or, in other words, have a limit to the number of people involved who could be interviewed. One should not critique qualitative research on the basis of quantitative methods. Instead, qualitative research should be evaluated based on four major areas: credibility, dependability, transferability, and promoting action and collaboration.

KEY CONCEPTS

attention to voice
autobiography
autoethnography
biography
bounded system
case study
catalytic authenticity (promoting action and collaboration)
credibility
dependability
emergent design
ethnographic research
external audit
gatekeepers
key informants
life stories
member checks
memoir
narrative inquiry
negative case analysis
oral history
peer debriefer
phenomenological research
saturation
subjectivity
transferability
triangulation

DISCUSSION QUESTIONS OR ACTIVITIES

1. What topics do you think could be examined using either a narrative inquiry or phenomenological study? How would you plan for such a study? What preparation would you need? How would you collect data for the study? How would you examine subjectivity or your own biases and assumptions?
2. Examine your own diaries, journals, photo albums, letters, and e-mails and consider how these might be used in a narrative inquiry about your own life. What would be the focus of the study? How might you carry it out?
3. Consider a lived experience that you and a friend have both had, such as attending summer camp, experiencing the death of a loved one, participating in a social justice activity. Interview each other about the experience. Try to identify commonalities and differences in your experiences.
4. Gaining entry into a school to conduct ethnographic research is often a complex process. Consider what specific issues you would review with the director of a private all-girls middle school where you are conducting research on peer relationships.
5. Discuss how the criteria for evaluating qualitative studies differ from those used in quantitative research and why different criteria are needed for each type of research. Select one qualitative study that you have read this semester and discuss how it did or did not provide strong evidence to support its

conclusions. What did you find convincing or not convincing when reading the results of the qualitative study?

SAMPLES OF QUALITATIVE RESEARCH STUDIES

Sample Narrative Inquiry Studies

Covin, J., & Dirkx, J. M. (2003). "Being called awake": The role of transformative learning in the lives of environmental activists. *Adult Education Quarterly, 53*(2), 99–118.

Davis Halifax, N. V., Gray, R., & Jadad, A. (2004). Self-portraits of illness: The patient's gift. *Canadian Medical Association Journal, 171*(7), 764–765.

Xu, S., Connelly, F. M., He, M. F., & Phillon, J. (2007). Immigrant students' experiences of schooling: A narrative inquiry theoretical framework. *Journal of Curriculum Studies, 39*(4), 399–422.

Sample Phenomenological Studies

McClelland, J. (1997). Knowing and being known: Parents' experiences with rural schools. *Journal of Research in Rural Education, 13*(2), 108–116.

Mills, L. J., & Daniluk, J. C. (2002). Her body speaks: The experience of dance therapy for women survivors of child sexual abuse. *Journal of Counseling and Development, 80*(1), 77–86.

Sample Ethnographic Research Studies

Markstrom, A., & Hallden, G. (2009). Children's strategies for agency in preschool. *Children and Society, 23*(2), 112–122.

Monzo, L., & Rueda, R. (2009). Passing for English fluent: Latino immigrant children masking language proficiency. *Anthropology and Education Quarterly, 40*(1), 20–40.

Sample Case Study Research Studies

Abril, C. (2009). Responding to culture in instrumental music programme: A teacher's journey. *Music Education Research, 11*(1), 77–91.

Lau, M. Y., Sieler, J. D., & Muyskens, P. (2006). Perspectives on the use of the problem-solving model from the viewpoint of a school psychologist, administrator, and teacher from a large midwest urban school district. *Psychology in the Schools, 43*(1), 117–127.

SUGGESTED READINGS

Clandinin, D. J, & Connelly, F. M. (2000). *Narrative inquiry: Experience and story in qualitative research*. San Francisco: Jossey-Bass.

Connelly, F. M., & Clandinin, D. J. (1999). *Storied identities: Storied landscapes.* New York: Teachers College Press.

LeCompte, M. D., & Schensul, J. J. (1999). *Designing and conducting ethnographic research: Ethnographer's toolkit* (Vol.1). Walnut Creek, CA: AltaMira.

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Wolcott, H. F. (1999). *Ethnography: A way of seeing.* Lanham, MD: AltaMira.

CHAPTER SEVEN

ORGANIZATION AND ANALYSIS OF QUALITATIVE DATA

CHAPTER OBJECTIVES

- Identify the processes involved in analyzing qualitative data, including data coding, description, identification of themes, hypothesis testing, and reporting and interpretation of data
- Discuss how a qualitative researcher goes about establishing themes and generating hypotheses
- Apply the process of qualitative data analysis to simple data samples

ANALYSIS OF QUALITATIVE DATA

Most qualitative researchers amass mounds of information in the course of their studies. For many reasons, qualitative researchers do not wait until all of the data have been collected to analyze them. They review their data as they are collected and record and write up their hunches, initial analyses, and questions in the form of research memos. As we noted in chapter 6, this is part of the emergent nature of qualitative research. Unlike quantitative research, in which data analysis comes after the study has been completed, in qualitative research, analysis of data occurs throughout the study and guides the ongoing process of data collection.

In this chapter, we illustrate the data from qualitative research and how they are analyzed. We first present a general description of the analysis of qualitative research and then discuss procedures most appropriate to each type of qualitative research. Examples of data from interviews and observations illustrate the analysis.

STEPS IN ANALYZING QUALITATIVE DATA

Data collection and analysis in qualitative research are inductive processes. As you may recall from previous chapters, this means that numerous small pieces of data are collected and gradually combined or related to form broader, more general descriptions and conclusions. Although the steps involved in qualitative data analysis vary according to the research questions asked and the type of approach taken, the steps listed here are common to most studies:

1. Prepare and organize the data.
2. Review and explore the data.
3. Code data into categories.
4. Construct thick descriptions of people, places, and activities.
5. Build themes and test hypotheses.
6. Report and interpret data.

Note that there may be some back-and-forth movement among the steps. One might need to review data before deciding how to organize it. The process of constructing descriptions of people or places might also lead one to reorganize some of the data. However, the steps tend to flow in this general direction.

Step 1: Prepare and Organize the Data

The first task in data analysis is to make sure that data are in a form that can be easily analyzed. Depending on the time and resources available, researchers may aim for different levels of depth in preparing their data. If interviews were tape recorded, preparation involves transferring the information from the recorded interviews into a written form. The quickest and least accurate approach involves making notes while listening to tapes from the interview and recording the general issues or ideas that are reported, using participants' own words as much as possible. Possible illustrative quotes are noted and recorded. This type of transcription involves data analysis as well as preparation, and it is not suitable for complex research questions or for beginning qualitative researchers. By analyzing the data at the same time that one is preparing and organizing it, the researcher's biases are more likely to influence the study's findings.

Most qualitative researchers prefer to separate the process of data preparation and analysis by transcribing interviews verbatim. Verbatim transcription is time consuming: typically a one-hour interview takes 6 to 8 hours to transcribe. The exact words of the participants are recorded, along with some aspects of nonverbal communication, such as pauses, laughter, interruptions, changes in vocal tone or emotion, and places where the tape is inaudible or not understandable. These nonverbal aspects of an interview are usually noted inside brackets, such as [laughter]. In group interviews, the person who is speaking is identified using initials or code names to guard confidentiality. Group interviews often involve several interviewers, one of whom records observations of interactions among group members and details that might not be captured on the tape. If the group observer records time intervals on the notes taken at the interview, notes from the verbal transcription and the observations can be examined side by side.

Other types of data preparation include the development or enlarging of photographs, labeling videotapes with identifying information (date, setting, or group), and making backup copies of data to store in a separate location (after months of data collection, you do not want to lose any data!). If computerized data analysis tools (discussed in more detail later in the chapter) are used, the data might also need to be put into a format compatible with the software program to be used.

Finally, the researcher must decide on a way to organize the large amount of data that a qualitative study typically yields. Data may be organized in many different ways, depending on the research questions and the method of qualitative research used. Following are some common methods for organizing data:

- *Site or location from which data were collected.* This method is common in studies in which multiple sites or locations were observed.
- *Person(s) or group studied.* Data may be organized by individual person or group, or data from persons or groups with similar characteristics or backgrounds might be grouped together.

- *Chronological order.* Data might be organized into the time periods in which it was collected.
- *Type of data.* Interview transcripts might be assembled together, separate from field notes and journals.
- *Type of event or issue addressed.* If interviews focused on different issues or observations of different events were made, the data pertaining to each issue or event might be grouped together.

Although the researcher needs to decide on an initial method for organizing the data, data may be reorganized after the initial analysis to look more closely at the categories and themes that emerge. For example, data from an ethnographic study might be reorganized by gender if this is an issue that emerges in the initial analysis.

Step 2: Review and Explore the Data

This step is a lot like jumping off the high-dive board at the swimming pool. A qualitative researcher might look with dread at the enormous pile of data waiting for analysis. However, all you can do is jump in and begin to explore by reading and looking through the various types of data collected. The initial review does not involve a careful reading for detail. Instead, one reads and examines data to get an overall sense of what is in them and whether enough data have been collected. Many researchers begin to jot down words and phrases that capture important aspects of the data in this initial review. However, the real purpose is to immerse oneself in the data and gain a sense of its possibilities.

Qualitative researchers often make notes to themselves about the data as they collect it; this means that they have already begun the process of reviewing the data before they begin their formal analysis of it. The initial review, however, is more comprehensive and involves examining all of the different sources of data together. Through the initial review, qualitative researchers seek to understand the scope of their data before they begin to divide them into more manageable chunks organized through codes. For many qualitative researchers, it is hard to say when the initial review stops and the coding begins, because one process leads naturally to the next.

All qualitative researchers engage in multiple readings of their data, with each reading serving a different purpose. They engage in an initial review to get a sense of the overall flow and structure of the data. In addition, data from interviews or conversations may be reviewed specifically to analyze the ways in which each individual uses language, noting words or phrases that seem distinctive or analyzing the audience that the participant seems to be addressing.

Step 3: Code Data Into Categories

Coding is the process of identifying different segments of the data that describe related phenomena and labeling these parts using broad category names. It is an inductive process of data analysis that involves examining many small pieces of information and abstracting a connection between them. Table 7.1 describes common code categories and examples of code names that might be used in analysis of an interview with J, a Latino youth discussed in Box 7.2.

As you can see, the codes describe general categories that can be used to organize the information contained in the data. There is nothing mysterious about coding. Here is a typical coding sequence.

1. Select an interview or set of field notes to review.
2. Review the data and think about ideas, behaviors, or other issues that seem important.
3. Highlight the part of the data (for example, segment of text or images) that relates to this idea and create a code word or phrase. Write the code in the margins.
4. Continue creating codes for the entire interview or field note.
5. Make a list of all codes created for this data set.

New codes are added as the researcher reviews other data sets. Most data sets use 30 to 40 codes initially, although complex studies might include more

TABLE 7.1 Common Code Categories and Examples of Code Names from a Life Story of a Latino Youth

Code Category	Code Names
Setting or context	Classroom, home, work, out of school
Personal characteristics	Self-esteem, skills, interests
Activities or actions	Hanging out with peers, band, classroom instruction
Events, experiences	School experiences, awards and recognitions
Perspectives of participants	Student perspective, parent perspective, teacher perspective
Feelings, emotions of participants	Pride, fear, love, anger, anxiety
Concepts, issues	Girlfriend's pregnancy
Relationships	Peer support, group differences, teacher-student relationships, parent-school relationships
Cultural context, social structure	Gender, ethnic identity, social class, race, ethnicity, immigrant status

than this. The process of coding can be conducted by hand or by computer. When coding is done by hand, the researcher writes the code in the margin of the data source (which was duplicated before coding began) and then organizes the data into piles with the same codes, cutting up data sheets as needed. When one has hundreds of pages of data, this can get messy. Most qualitative researchers reserve a part of their office or an entire room for data analysis, and one researcher we know who conducts extensive qualitative research even built an addition on her house for her research work!

Any given data segment might be viewed differently by two different researchers or even coded using more than one label by a given researcher. Therefore, qualitative researchers continually read, reread, and reexamine all of their data to make sure that they have not missed something or coded it in a way that is inappropriate to the experiences of the participants. You can try this out with a classmate. We have already created some codes for you in Table 7.1 (for example, school experiences). Select several code names from Table 7.1 and see if you both can apply these to the interview transcript in Box 7.1. Compare your coded transcripts with your classmate to see if you interpreted the data in similar ways. Also, consider if there are areas for which you need to create a new code name.

Several computer programs are available to assist in the analysis and especially the coding of data, although we emphasize that the computer is only a tool and the researcher still makes decisions about how to do the analysis and what the results mean. However, computer programs such as NUD*IST (Non-numerical Unstructured Data Indexing, Searching, and Theorizing) or Ethnograph can be used to combine files from different data sources, select and code the data, and organize data using codes or hypermedia links. Computer programs facilitate the processes of applying multiple codes to the same segment, viewing data using different organizations of codes, and displaying data using different formats. (For more in-depth information on using computer software for qualitative analysis, see Judy Norris's QualPage Web site at http://www.qualitativeresearch.uga.edu/QualPage/qdaresources.htm.)

Whether the researcher uses a computer program or manual cutting and pasting, coding involves a process of continual refinement and abstraction from the data. The initial 30 to 40 codes are gradually combined and reduced to about 15 to 20, with the goal of eliminating overlap and producing a more coherent view of the patterns in the data. For example, in the analysis of J's data researchers decided to combine the codes for school experiences, awards and recognitions into one code that they called positive school experiences. When the researcher is satisfied that the major ideas and issues in the data have been identified through codes, the next step is to use the codes to organize data and construct descriptions of the data.

Step 4: Construct Thick Descriptions of People, Places, and Activities

Another step in qualitative data analysis is the writing of **thick descriptions**—detailed descriptions of the people, places, and events in the study. The goal is to provide rich, in-depth descriptions of the experiences, perspectives, and physical settings represented in the data. You may recall that we discussed thick description of field notes in Chapters Two and Five. Descriptions in data analysis often involve expanding on one's field notes and combining notes and interviews with the same codes into more integrated descriptions of people, situations, and places. These may be written during either the data collection or the data analysis part of the study. Box 7.2 presents a sample of a thick description of the Latino youth, J, described earlier. Note that the description captures the details of the setting and the actions of the characters, both of which are considered critical to portrayals of an experience. Thick descriptions must be situated in a particular context in order to fully reveal their meaning.

Writing good detailed descriptions of even the most ordinary aspects of everyday life is an essential part of qualitative research. Good qualitative research can make readers feel like they are living the experiences described. This is the real power of qualitative research. Narrative researchers use the term **restorying** to refer to the rewriting of the participant's story in the researcher's words. This term distinguishes between the participant's original story and the researcher's selection of key elements from the data and organization of these elements into a sequence. As noted in chapter 6, good qualitative researchers collaborate closely with participants and use techniques such as member checks to ensure that the story accurately represents the participants' experiences.

Step 5: Build Themes and Test Hypotheses

Coding and description comprise the first two levels of qualitative data analysis. Deeper analysis in which explanation of the events and issues represented in the data occurs as the researcher continues the process of abstraction by identifying major and minor themes in the coded data. **Themes** are typically "big ideas" that combine several codes in a way that allows the researcher to examine the qualitative subquestions guiding the research. In other words, themes provide the organizing ideas that the researchers use to explain what they have learned from the study. Like codes, themes are usually described in a few words or phrases, but they identify the major concepts or issues that the researcher uses to interpret and explain the data. The goal is to reduce the number of codes and identify seven to eight themes that accurately describe the data. The researcher then reexamines subquestions and the data using the themes as organizational frameworks to see if this results in a deeper understanding of the data.

BOX 7.1
Sample Interview Transcript from a Narrative Inquiry Study

J is an 18 year old high school student in a single parent home who is a teenage father. A series of interviews was conducted with this youth to elicit his perspective on the important people and events in his life. Below is an excerpt in which he discussed how he discovered his talent for music.

Interview 1 with Youth

J: the class I really, really like out of all of them is probably Band. Because since grammar school, I remember one time I had a teacher, Mr. Bradley O'Brien. He's teaching at Lane High School now, he's the orchestra conductor. But he used to be our only music teacher in school and . . . Since I play percussion, he is actually the person who found my talent. He is the person I owe it all to because one day he took a snare drum to our class—he usually takes different instruments, but one day he took a little drum and he said, "I want to see if anybody can try to do a drum roll." So then he wanted volunteers, he was like does anybody else want to try? A girl went before me. So I was like, "yes, sure I'll try." So he just did the drum roll and I just did it exactly like he did it. And he was like "Wow, you know, have you ever thought of joining the band?" And I was like, "no, I never thought of it really. I didn't even know we had a band." So he was like, "okay, just think about it and see if you want to give it a try." So I was like, "Fine, I'll think about it." I was just saying I'll think about it. I wasn't really too interested because I was like, yeah, how would it be, and this and that? So he bothered me for like a whole week, saying, "Do you want to join the band? Do you want to join the band?" So I told my mom, "I'm thinking about joining the band." And she was like, "well, you know if it's going to help out with your education, and anything like that. You know how I always support you and stuff." So I was like, "yeah, okay." So I told the teacher, "Yes, I will join." So freshman year I had to take either art or music so I first took art, so that later on I could just take music from there. So from sophomore year I started with beginning band here and they usually take you to intermediate but I went straight from beginning to advanced.

I: Wow.

J: so I'm now, my first year, that was my sophomore year beginning band and I went to advanced then, but I wasn't section leader. Then next year I was section leader, I am also section leader now. And from there, Mr. A, our band conductor, he gave me a lot of opportunities. He made me go to all city band which is for Chicago, I played with them for two years. I've been in the Ravinia Jazz Festival, which is a Jazz college program and they offer a lot of opportunities too. They also gave me a scholarship to go to a Jazz camp in Louisville KY? For the summer. And it was a nice experience, you know, it's for a week. They paid for the travel, food, all accommodations and stuff like that too.

I: oh, that's great!

J: so that was pretty cool! And right now when I go to <name of state college>, I'm going to be playing in their band also. They gave me a scholarship like I told you earlier. You know they gave me a scholarship and I'm going to be in their band. I can also major in something else, I can study something while I'm playing in the band. And I'm also going to be playing in the Jazz band and the Salsa band.

. . . so it's like I'm taking music from just little, you know, just volunteering to playing Jazz with my class, I was taken to all these different places, and you know, that I never actually thought that I could make it. So I'm glad that happened.

Interview 2 with Teacher

One of the youth's teachers from his Journalism class was also interviewed to get his perspective on the youth's life.

Interviewer: You mentioned that J gets along with everybody, so can you give me an example of how you see that in your class?

Teacher: I think [pauses as train goes by]. I think J could get to know the engineer on that train. [laughter]

Interviewer: Yeah, right.

Teacher: I think one of the, one of the best examples I have of J being able to talk to anybody and actually helping people who are really shy to come out of their shyness is, there is a student in the journalism class, M, who, umm [laughs], this is probably, probably not a good story. At the beginning of the year, I was trying to figure out why this kid wouldn't do anything. M was a big, big kid, a little older than most of them. And he just did nothing the first five weeks of school. So I looked up who his counselor was, and I went to the counselor and said, "I have a student, and I have real big problems getting him to do anything in my class." And she asked who the kid was, and I told her M. And she went [with exaggerated horror], "Oh, no! You don't have him in journalism, do you?" I said, "Yeah." She's like, "Oh, oh my God! You know some of the kids we say that they are very special. This one is just retarded." [Teacher shakes his head and looks at Interviewer with disbelief and disgust.]

This is coming from the kid's counselor! So she actually tried to get him out of the journalism class, but once I got the information that M has a severe learning disability I gave the kid things that he could do, and he was fine. And for a good part of the year, he would just quietly do those things. And I was the only person that he would talk to. But then J just took it on himself to start talking to M. When J was talking with a group, he would turn around and ask M a question. He would frequently invite him to help him edit his stories or ask him about ways to illustrate his stories. And now M just walks around the room and talks to everybody.

BOX 7.2
Thick Description of a Latino Youth

Jaime moves with grace through the crowded halls of his high school, pausing to laugh with a friend or clasp hands in greeting. His progress through the hall is slow but smiles lay in his wake because Jaime seems to elicit good cheer in everyone that he greets. Somehow he connects with each friend in his path as he navigates the 1200 feet between his 4th and 5th period classes in this large, urban high school and arrives just seconds after the bell signaling the end of the 4 minute interval between classes. He flashes a final smile at the teacher who is greeting students at the door and slides into his desk.

Jaime has found his place at this school and is so eager to learn that he has decided to stay one more semester although he has more than enough credits to graduate. He is taking elective classes such as Journalism and Band, to refine his research and writing skills and pursue his love of music. He is also taking Italian, his fourth language after Spanish, English, and French. In the fall he will study computer science at the state college supported by a scholarship to play percussion in the college's jazz and salsa bands. He dreams of playing for a large city symphony orchestra, but plans to study computers to ensure that he has the skills to "better himself" and support his son and girlfriend/fiancée.

Jaime has accomplished much in his 18 years of life, and he is aware of how fortuitous events and caring people have contributed to his successes. He remembers his early childhood in Mexico, growing up in Iguala in the state of Guerrero, as a time of play, warm family relationships, and early education. Family members recall how he loved jumping into the pristine river near his home, fully clothed. The family did not have a lot of money, but Jaime treasures memories of that time, such as, when his cousins gave him his first bicycle (Lodico & Voegtle, 2005, pp. 137–138).

For example, the data presented earlier on J emerged from a study that included the following subquestion:

> *What types of personal characteristics and experiences lead to positive development in low income youth?*

To identify a theme we first reexamine the code words in Table 7.1. We then select several code words related to the above subquestion. We selected school experiences, skills, and self-esteem. What might be an overall theme that would combine these codes? We thought that emerging resiliency could be a theme uniting these codes, because J's positive school experiences, his musical and social skills, and his high self-esteem all contribute to his resiliency.

Reexamination of the data using the themes may lead to tentative explanations of the events in the study. These tentative explanations are **hypotheses** about the issues identified in the qualitative subquestions. Unlike research hypotheses used in quantitative research, hypotheses in qualitative research emerge from the data analysis. They are not predictions about what will happen; rather, they are tentative explanations of the processes underlying what has been observed. Once a hypothesis has been formulated, data are reexamined for both confirming and disconfirming evidence. Confirming evidence is often obtained through **triangulation**, the process of comparing different sources of data (for example, interviews and observations) or perspectives of different participants. One hypothesis regarding J might be:

Positive school experiences promoted resiliency and self-esteem in J.

We next review the data to find further evidence to support this hypothesis. We thought that J's descriptions of his supportive teachers and the opportunities provided by his school supported the hypothesis. Examine the data in Box 7.2 to see if you find any other evidence from his teacher to support or discredit the hypothesis.

Given the nature of qualitative research, there are times when participants offer conflicting perspectives. When conflicting perspectives are found, qualitative researchers must reexamine other data sources to see if the differences can be resolved. In some cases, if the differences cannot be resolved, the researcher may decide to simply present the different perspectives or identify the differences as something unique to a certain individual or group.

There were no conflicting perspectives in J's case, so we instead reexamine the case study involving Max described in Chapter Five. The researchers found that Max's teacher and mother disagreed regarding his readiness for kindergarten. Compare the descriptions of Max by his mother and his teacher in Box 7.3. What would you conclude regarding Max's readiness for kindergarten?

In the final write-up of the case study of Max, the researchers simply described the differences in the mother's and teacher's views regarding his kindergarten readiness. They later followed up to see how Max was doing in kindergarten. Interestingly, although he still had some problems with peers, he seemed to thrive in the more structured kindergarten setting. So it seems both Mom and the teacher were right!

BOX 7.3
Contrasting Perspectives in a Case Study

Mother's perspective: When talking about Max's social skills, Mom has some real concerns. She says that his "social skills are a little lacking with his peers." She believes that Max has difficulty initiating play activities with other children. He appears to "be rigid, and I just wish he would be more flexible," says Mom. He does not appear to want to compromise with his peers and wants his own way. Mom says that Max does have one good friend at school with whom he shares similar interests. They apparently play pretty well together until the friend does something that Max doesn't want him to do. For the most part, Mom believes that Max is "kind of a loner, unless kids want to do what he wants to do." In the future, she hopes that Max will learn to compromise. She worries that he will be the kid who won't sit in his chair and is always in the principal's office because he wants to do what he wants to do. Mom feels very strongly that Max needs a structured classroom. "He needs to have a time to do one thing and a time to move on to something else." Mom does not believe that Max can get such a structured environment in a preschool setting. Given the fact that academically Max is so far ahead of his peers, Mom has been looking at different kindergarten programs despite the fact that his preschool teacher believes that Max is not ready to move on to a kindergarten classroom and needs one more year in preschool.

Teacher's perspective: Ms. Smith is particularly concerned with Max's tendency to "get into the faces" of other kids. He knows "exactly whose face to get into." She says that while she has tried various techniques to eliminate the behavior, nothing has seemed to work. She praises him when he is playing cooperatively with the other children. She has tried a time-out when he is behaving inappropriately and verbally tells him that his behavior hurts others' feelings. She explicitly tries to tie the "in-your-face" behavior to the hurt feelings of others. Ms. Smith says that Max does not seem to care and that the reprimand does not appear to make any difference to him. While Ms. Smith believes that Max has made tremendous progress over the first several months of this year, she is concerned that socially he might not be ready for kindergarten. She believes that Max is clearly ready to do the academic work of kindergarten, but his social skills need to be further developed. She has recommended another year of nursery school, especially given the fact that he is just turning four.

Both of the cases presented examined single persons. Some researchers might compare the experiences of several people in one study and look for variations in participants' experiences that reveal unique aspects of their individual experiences. Kennedy and Voegtle (2000) interviewed nine persons who were considered to be allies of gay and lesbian persons. They found that all allies developed strong ideological frameworks that emphasized acceptance of differences, although these frameworks differed for different allies. Excerpts from interviews for three allies follow. How would you describe the common themes underlying all allies? How do their ideological frameworks differ?

> *Ally 1:* I feel it really does fit with my politics and always has. As someone who related more to republican side of things mostly because of the idea of freedom. That aspect is important to allow people to be individuals and be themselves and to not permit people from being able to accomplish things and doing things in life because of certain things they can't help or something that is inherent within. . . . To me I don't see gays getting married as a special right. It is the same right that I had; I had the right to marry the person I loved. And I think everyone should have the right to do that.
>
> *Ally 2:* An ally is somebody who acts against racism in support of people of color. An anti-racist is somebody whose position and actions in politics are to act against racism in any situation where there are any people of color there or not. Whether you are in a group or not, it doesn't have to be directly in support of people of color. I am an anti-racist wherever I am. Whether there are people of color there or not. So parallel to that I've attempted to develop a position like that around homophobia. That it's not enough to be an ally. That anti-homophobia, anti-heterosexism is part of my politics at my work.
>
> *Ally 3:* Knowing and caring about people who were victims. And not being able to stand by and watch that without saying something. Feeling the incredible loss of love. Most of this has been in context of the church. Feeling the incredible loss of the church from excluding these men and women with incredible gifts and insights, pains and joys to share with the rest of us. To miss out on all of that is just such a shame. And so that's what's pushed me. . . . It's solidly based in my understanding of what it is to be a Christian and solidly based in the way I read scripture and it's solidly based in my understanding of what it means to be a person of faith. . . . Everything, everything in me, everything that I can understand in scripture, everything that I can understand in who I think God is, everything who I can understand who Jesus is says we need to be welcoming and affirming (Kennedy & Voegtle, 2000, ¶39–41).

As qualitative researchers formulate themes and hypotheses, they often probe beneath the surface detail of their data. Connelly and Clandinin (1990) discuss how researchers' retelling of a participant's story might broaden the narrative by relating a specific action to a "person's character, values, way of life or perhaps the social and intellectual climate of the times" (p. 11). At other times, the researcher might "burrow" more deeply into an event by examining its emotional impact on a participant or hypothesizing about the origins of the behaviors described. These broadening and burrowing processes often result in positive growth and self-understanding for the participants, a result that fits well with the advocacy/liberatory goals of some qualitative researchers.

Qualitative researchers also engage in deep analysis of their own subjective experiences in conducting research and analyzing the data. Narrative researchers sometimes provide a separate narration of the researcher's story along with the narrative describing the participant's story. While all qualitative researchers examine their own biases, the intense intimacy of both narrative and phenomenological research are often life changing for both the researcher and participants, and documentation of this process forms part of the study.

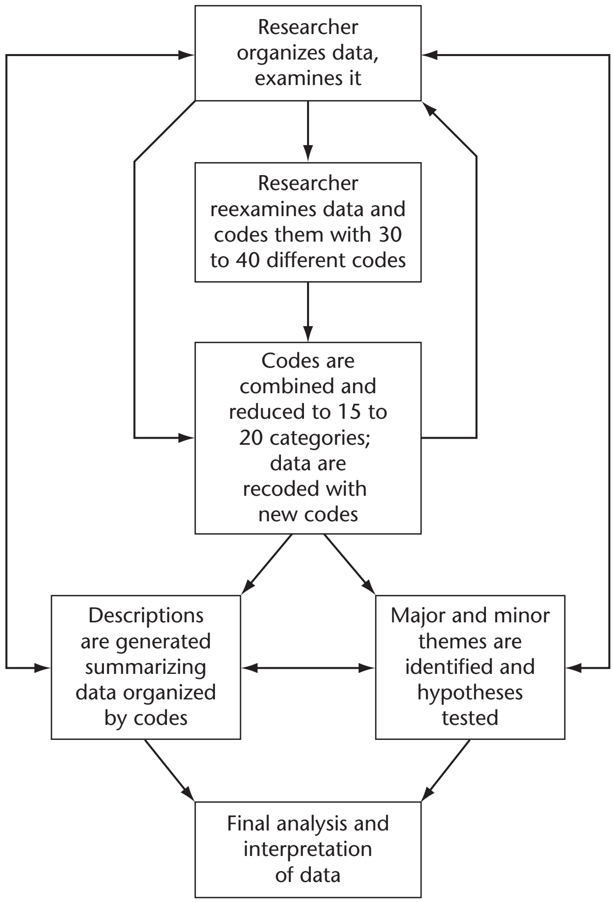
Although we have described the processes of coding, description, theme building, and hypothesis testing through review of data as if they are separate steps, qualitative data analysis is an iterative process: the steps are repeated several times, with something new added in each iteration. The process is repeated until the researcher feels that the research questions have been answered and sufficient meaning extracted from the data. Figure 7.1 displays the iterative nature of the processes of coding, description, and theme identification.

Step 6: Report and Interpret Data

The final step in qualitative data analysis is the writing of the research report, including the researcher's interpretations of what the data mean. Most qualitative research is reported in a mostly nonquantitative, narrative manner, which often makes it enjoyable to read. The research report may be organized using any of several different **report formats** summarized in Table 7.2. The choice of format may be determined both by the results of the data analysis and by the researcher's philosophical framework and purpose in conducting the research. Many researchers include visual diagrams or images to represent the complex array of events, issues, or themes that emerged from their data analysis. Dramatic presentations are also a common vehicle for reporting results. For example, Spaulding, Lodico, Jones, and Gligora (2004) presented photographs taken by the youth in a summer camp academic enrichment program to illustrate the major findings of their research evaluating the program. A more dramatic example is how Eve Ensler transformed her interviews with women who had been victims of sexual violence into a play, *The Vagina Monologues* (1998) that has been performed in hundreds of locations.

Reports of qualitative studies usually include extensive samples of quotes from participants. By using the participants' own words, researchers aim to build the reader's confidence that they are accurately representing the reality of the

FIGURE 7.1 Steps Involved in Analyzing Qualitative Data



persons and situation studied. Some narrative researchers put extended participant quotes in a separate and uninterrupted section from their analysis. As in quantitative research, qualitative studies also report the methods used to collect and analyze data, but the criteria used to evaluate qualitative studies differ.

It is often difficult to distinguish in a qualitative study between the reporting of findings and the interpretation of the findings. According to Lincoln and Guba (1985), interpreting qualitative data involves making sense of the "lessons learned" by looking for their larger meaning. Interpretation might involve relating the findings to previous published studies or to a theoretical framework. For example, Kennedy and Voegtle (2000) related their findings on ally development to

TABLE 7.2 Report Formats for Presenting Findings from Qualitative Research

Format	Description
Alternative or performance based	Text is presented using a performance-based format such as a story, song, dramatic performance, or highly personalized account called an autoethnography. This format is useful in capturing the intense emotionality of a setting or experience.
Amalgamation	Researcher analyzes data from several people and creates descriptive portraits of the "types" of persons involved in the study. Each portrait is based on multiple persons so that it protects confidentiality of the information. Activities may also be amalgamated into a "typical day or week."
Natural history	Text structure parallels or "recreates" the process of exploration and discovery that occurred during fieldwork. This format conveys a strong sense of the people, the setting, and the interactions involved in the research, although it makes it difficult to do theme analysis.
Theoretical	Text is organized around a theory used throughout the report. A developed theory may serve as the framework for reviewing literature and collecting data, as in a theoretically oriented case study. Grounded-theory approaches organize writing in terms of the creation of a new theory that explains the data or the modification of an existing theory based on the data.
Traditional scientific	Text is presented in the traditional style of research reports including Introduction, Review of Literature, Method, Results, and Discussion sections. Results and Discussion sections include analysis of themes.

Sources: Adapted from Creswell (2005), Denzin and Lincoln (2005), and Glesne (2006).

Allport's contact theory of racial prejudice (1955). Some qualitative researchers provide more extensive discussion of research literature and theoretical frameworks after their data analysis in the interpretation section of their reports. Interpretation of qualitative data may also involve final personal reflections by the researcher because the researcher has typically invested considerable time and emotional energy in collecting and analyzing the data. Researchers taking an advocacy or liberatory approach discuss the implications for taking action suggested by the data analysis, especially if the data indicate that certain groups are being treated unfairly. For example, Kozol (2000) has engaged in extensive political work to advocate for policy changes and resource allocations to benefit urban school districts. Finally, as in quantitative studies, interpretation of the data in a qualitative study may also include discussion of the limitations of the study and ideas for future research.

SUMMARY

Data collection and analysis in qualitative research are inductive processes. Like quantitative studies, qualitative studies use certain steps that the researcher has to conduct. The first step is preparation and organization of the data. Several levels of data collection from interviews are available, ranging from taking notes from tape-recorded interviews to making full transcriptions from such tapes. Qualitative researchers try to keep data collection and preparation separate from data analysis. Data can be organized in several ways, for example, by site or location, person or groups being studied, or chronologically. The second step is the initial review and exploration of the data. As part of this process, the researcher initially reads through all data to get an overall sense of what are in the data and whether enough data have been collected. The third step is the coding of data into categories. To do this, qualitative researchers typically continually read, reread, and reexamine all of their data to make sure that they have not missed something or coded them in a way that is inappropriate to the experiences of the participants. Qualitative researchers use 30 to 40 codes, although complex studies may require more. Computer programs such as NUD*IST or Ethnograph are used by qualitative researchers. Following this, the next step is constructing descriptions of people, places, and activities. Once the data have been coded, the researcher writes detailed descriptions of the people, places, and events in the study for the purpose of providing rich, in-depth descriptions, often referred to as thick descriptions. Building themes and testing hypotheses comprise the next step. Themes are big ideas that combine several codes in a way that allows the researcher to examine subquestions guiding the research. Researchers seek out evidence that might disconfirm their hypothesis by continually reviewing their data or collecting new data to examine their hypothesis. Reporting and interpreting data are the last steps in the qualitative research process. A wide variety of report formats may be used.

KEY CONCEPTS

- coding
- hypotheses
- report formats
- restorying
- themes
- thick description
- triangulation

DISCUSSION QUESTIONS OR ACTIVITIES

1. Select a movie that depicts classroom interactions between a new teacher and students (for example, *Stand and Deliver*). With other students in your class, identify three research questions that you can use to guide qualitative data collection. View the film and record field notes on what you observe. Combine the field notes with those of other students, and identify common themes that you see. Based on your initial themes, formulate a hypothesis that could be examined through further data collection.
2. Find a published qualitative study that interests you. Discuss with a classmate what type of report format was used and how the study might have been presented using a different report format in Table 7.2.

SUGGESTED READINGS

Merriam, S. B. (1998). *Qualitative research and case study applications in education.* San Francisco: Jossey-Bass.

Morse, J. M., & Richards, L. (2005). *README FIRST for a user's guide to qualitative methods*. Thousand Oaks, CA: Sage.

Schensul, J. J. (1999). *Analyzing and interpreting ethnographic data* (*Ethnographer's Toolkit,* Vol. 5). Lanham, MD: AltaMira.

CHAPTER EIGHT

DESCRIPTIVE SURVEY RESEARCH

Quantitative Research

CHAPTER OBJECTIVES

- Define the purpose and characteristics of descriptive survey research
- Distinguish among the different types of descriptive survey research
- Outline the seven steps to carry out descriptive survey research and describe the essential elements of each
- Distinguish between initial research questions and subquestions
- Distinguish between survey sections and individual survey items
- Know the criteria for evaluating descriptive survey research studies

RESEARCH VIGNETTE

An intern at a local high school, Alysia has been asked by the school principal to determine the students' perceptions of the school's new dress code that will be implemented with the new school year. Alysia is taking an educational research course as part of her educational administration certificate program. She consults with her professor and decides that the best way to study perceptions is to conduct a descriptive survey research study. The high school is made up of grades 9 through 12, and the total school population is 2,500. She decides to randomly select 50% of the students. She constructs a survey with 20 response items in which students have to circle a number on a scale of 1 through 5 indicating the degree to which they agree with the statement (a Likert scale). She includes three open-ended questions to allow the students an opportunity to express any further feelings or opinions. Before she sends out the survey, she gives the survey to a group of students in their study hall and asks them to critique the survey. Based upon their feedback, she constructs the final survey and mails copies to the random sample with a self-addressed stamped envelope (SASE). In her cover letter, she asks that the surveys be returned to her within 3 weeks. Three weeks later she has received 20% of the surveys back.

CHARACTERISTICS OF DESCRIPTIVE SURVEY RESEARCH

Survey research is everywhere. Take, for example, the last time you sat down to enjoy a meal with your family or friends after a long week of graduate school, and the phone rang right as the meal was set on the table. You answered it to find a telemarketer on the other end explaining that she was conducting a brief survey and would like to ask you some questions about the radio station you listen to on your way into work each morning. And as you are about to hang up, she adds, "And it will only take a second of your time." Famous last words, right?

Whether someone comes up to you in the shopping mall with a clipboard, calls you on the phone, or sends you a paper survey in the mail (or via e-mail), it is an example of descriptive survey research. The approaches share the following common characteristics:

- An instrument has most likely been developed by the researcher.
- The majority of responses to the questions on the survey are quantitative (for example, ratings) or are summarized using numbers (for example, number of people choosing a given answer).
- Typically, the sample is selected from a larger population or group to allow the study's findings to be generalized back to the larger group.

The various approaches to survey research all have the same purpose: gathering opinions, beliefs, or perceptions about a current issue from a large group of people. In educational research, these issues can be wide ranging and may include, but are certainly not limited to, high-stakes testing, parent involvement, school improvement, classroom instructional practices, behavior management techniques, and after-school or summer enrichment programming.

Think about how Alysia might carry out her study. How would her study illustrate the characteristics of descriptive survey research? There are different types of descriptive survey research. By altering the pattern of survey administration and sampling, different types of studies can be conducted to allow a wider array of research questions to be answered. In other words, by changing *when* and *to whom* the surveys are administered, different kinds of research questions can be asked and answered. Presented following are several different, commonly used survey designs along with examples of research questions that would help guide their development. We first discuss the types of descriptive survey research and then the steps for conducting descriptive survey research.

One-Shot Survey Design

If you have been involved in a survey study, you probably used the most common design, known as the **one-shot survey design**. This approach is pretty much what it sounds like: surveys are mailed to participants at one point in time to gather their perceptions about a current issue.

Longitudinal Survey Study

In a **longitudinal survey study** a sample is selected and the entire sample is sampled periodically. The main function of the longitudinal study is to "track" participants over an *extended* period of time. It would not be unusual for a longitudinal study to cover 10, 20, or 30 years. The purpose for collecting data over this long period of time is to investigate how participant perceptions or attitudes about an issue change (sometimes quite dramatically) over many years, and to follow participants as their lives progress and their interests and philosophies change.

Trend Survey Study

Trend studies are typically used to examine the perceptions of groups that are having or have had a shared experience at a particular point in time. For example, a researcher may want to ask new teachers fresh out of their preservice teacher education programs how prepared they feel to teach. Surveys are administered to new graduates each year for five years. In this situation, the researcher is *only* interested in what *each group* of *new teachers* think (maybe capturing some of their unrealistic expectations or naiveté). The researcher is **not** interested in documenting initial perceptions and then showing how that same group's perceptions changed over time; rather, the researcher wants to describe the perceptions of each new group of teachers as they graduate and to document any trends that emerge. For example, would new teachers graduating in different years have different perceptions about the value of courses that focused on content versus pedagogy? Examples of national trend studies include:

- The Centers for Disease Control's Youth Risk Behavior Survey
  www.cdc.gov/HealthyYouth/yrbs/index.htm
- The University of Michigan's Monitoring the Future Survey on adolescent drug use
  www.monitoringthefuture.org

Follow-Up Survey Study

Let's say the same researcher was interested in knowing whether these new teachers stayed in the teaching field or found employment in other areas. In a **follow-up study,** these new teachers are surveyed upon graduation and again the following year to see how they are doing. In some studies follow-up may be repeated for a couple of years. Some of the teachers may have left the field during the time when no survey was administered. A follow-up study would help the researcher document the rate at which teachers in the study are leaving the field.

Cohort Survey Study

For this design, the researcher uses the same population each year but selects different samples from that group over time. For example, a sample of new teachers entering the field would be surveyed the first year. The next year, the same population of teachers who were selected the prior year (and now completing their first year of teaching) would be used, but another sample would be drawn from the original population. This process would repeat itself again and again for a number of years, creating a rigorous **cohort survey design**. Keep in mind that a key step in preparation for this design is to identify a very large initial population so that the researcher can draw repeatedly from the group and not survey the same person more than once.

Descriptive-Comparative Study

In some descriptive studies, participants are asked to report on their own activities, behaviors, and feelings using brief reporting forms. For example, Larson

(1989) examined the activities of adolescents by having them report what they were doing at randomly selected times during the day. Larson used a unique sampling procedure in which adolescents completed a written report on the activities they were engaged in when the beepers they had been provided went off. The researchers did not try to change the activities of the adolescents but simply described what the adolescents were doing. The results were summarized showing the percentage of time engaged in each activity. Larson did examine age and gender differences in the reported activities. This type of study is sometimes referred to as **descriptive comparative,** because it describes differences between groups but does not try to explain why these differences occur. The types of group differences that are explored in descriptive comparative studies often involve demographic variables such as age, ethnic group, gender, grade level, or job position.

The differences among these types of survey designs are summarized in Table 8.1.

STEPS IN CONDUCTING DESCRIPTIVE SURVEY RESEARCH

While there are different types of descriptive survey research, steps that one would take to conduct any type of survey study are similar. You will find that some of the steps presented here apply to other quantitative studies presented in later chapters.

Step 1: Identify a Research Topic

Since descriptive survey research typically describes attitudes, beliefs, behaviors, and perceptions, topics typically follow current trends and issues. In educational research, these issues often parallel changes in curriculum, instruction, assessment, leadership and policy, to name a few. Since these issues in education are often subject to change, descriptive survey research allows us to gain ongoing understanding about the way people think about these topics or what they do.

Step 2: Conduct a Review of the Literature

As in any research study, researchers first examine past literature to extend their current knowledge on their topic. In descriptive survey research, the literature is examined to learn what questions have been examined, what groups have been studied and what instruments have been developed and used. The review of literature often helps the researcher to develop the survey and design the study.

TABLE 8.1 Types of Descriptive Survey Designs

Type of Study	Purpose	Sampling	Number of Surveys and Administrations
One-Shot Survey Design	Determine current perceptions/ behaviors of one group at one point in time	Ideally, one random sample is selected from a larger population	One survey One administration
Follow-Up Survey Study	Determine current perceptions/ behaviors of one group and possible changes at one later point in time	One random sample is selected and the same group is surveyed at two different points in time	One group Initial and follow-up survey (which may differ) Administered at least two points in time
Longitudinal Survey Study	Follow one group for a long period of time to see how perceptions/ behaviors may change or stay the same	One sample is selected and surveyed many times for a lengthy period	One group Surveys may include some of the same items and some new ones to determine changes
Trend Survey Study	Examine trends in populations that are new each year but have similar experiences at different times	Each year a **new sample** is selected from a **new population** identified as having similar experiences to past populations	Same survey administered each year May be administered each year for long periods of time
Cohort Survey Study	Examine perceptions/ behaviors over time in the same population using different samples from that population	Each year a **new sample** is selected from the **same population**	Same survey administered each year May be administered each year with new sample as often as needed
Descriptive Comparative Study	Examine differences in perceptions/ behaviors without trying to infer causes of difference	One random sample is selected, and demographic information is used to subdivide the sample into smaller groups	One survey Administered once Results are analyzed for group differences

Step 3: Develop Research Questions

Although descriptive survey research is a type of quantitative research, it begins, as in qualitative research, with an initial research question and a set of subquestions. These research questions are based upon the information gathered by the researcher from the review of literature. The following are examples of initial research questions:

- What do elementary and middle-level teachers believe are the main benefits of and/or barriers to integrating technology into their instruction?
- What do high school administrators perceive to be the issues surrounding school safety?
- What do parents of elementary students believe are ways to become more involved in school and school-related activities?

Notice how these initial questions illustrate the characteristics of descriptive survey research. First, in each example, *who* is being surveyed (the sample) is clearly specified: elementary and middle-level teachers, high school administrators, and parents of elementary students. Second, all questions include verbs or "action" words that inquire about people's perceptions of their thoughts, feelings, and actions. Third, the issue being investigated is clearly defined: technology integration, school safety, and parent involvement in school activities.

As you examine the vignette, what do you think Alysia's initial research question was for her study? It is likely that her question was something like the following: *What are student perceptions of the new uniform policy?*

After generating an initial research question, students are often perplexed by how they can "fill up" an entire survey with questions based on the one research question they have created. The answer is, they cannot. The initial question should be thought of as a broad research question, and although every survey study needs to begin with one, the researcher must work to develop more specific questions or subquestions. Consider Alysia's initial research question: *What are student perceptions of the new uniform policy?*

Examples of subquestions that would support this initial question include

- How do students believe that school uniforms will affect the sense of community at the school?
- How do students believe that school uniforms will affect their academic performance?
- What effect, if any, do students believe that school uniforms will have on the safety of their school?

Although these subquestions are very specific, they are very much aligned with the initial research question. Each subquestion "chips away" at some aspect of school uniforms from different perspectives. In a descriptive survey research proposal, like the one you may write for this course, these subquestions are listed at the end of the review of literature and serve as a guide for writing the sections of the survey. Since Alysia has three subquestions she will have three sections in her survey designed to answer those questions. Aside from the demographics she collects, she will have one section on school community, one on academics and one on school safety. The subquestions are *not* quantitative in nature. It is in the detailed survey questions (which we discuss later in this chapter) that the researcher provides the structure that allows participants to provide quantitative responses. Creating subquestions is an ongoing process; the more you read and learn, the more you will want to go back and refine your subquestions. In fact, researchers commonly go back and make revisions to these subquestions while developing their surveys.

Step 4: Develop the Survey

A **survey** or **questionnaire** is the main tool or instrument used to collect data in a descriptive survey research study. Since survey researchers typically study issues and behaviors that change over time, they usually develop new instruments or refine existing instruments as they go. A common misperception about survey development is that it is easy—just put some questions down on a piece of paper and—"Voilà!" you have a survey ready to mail out to participants. This could not be further from the truth. Although it may appear to be easy, survey design and development requires a significant amount of knowledge, planning, and skill to execute correctly.

Sections of a Survey Presented next is an overview of the sections of a survey. Keep in mind that not all surveys are the same, but in general, most surveys have all or most of these components.

Instructions and Cover Letter The instructions or directions that the researcher provides to help guide the participant to complete the survey or questionnaire are some of the most important (and difficult) aspects of survey development. The researcher wants to make sure the surveys are answered accurately and returned, so directions and a statement of the purpose of the study need to be the first thing a participant sees. A cover letter is one method that researchers commonly use to relay such important information to participants.

Cover letters tend to work best for surveys that are being sent to participants through the mail. For situations when the researcher is collecting data from a site, for example, from a large group of teachers attending professional development training in technology literacy, cover letters are unnecessary. In those situations, it is appropriate for the researcher to read aloud to the group the directions and other information that would appear on such a cover letter as part of an introductory statement. If a cover letter is not used, a paragraph that appears at the very top of the survey can contain the essential information needed (see Exhibit 8.1 for a sample of what Alysia might write for her directions).

EXHIBIT 8.1
Student Perception Survey

PLEASE RETURN by July 30

You have received this survey as part of a research study. The purpose of this study is to investigate student perceptions of school uniforms. This survey was designed to gather your beliefs about uniforms and how you believe uniforms will influence your school. Your responses are confidential and will not be shared with anyone in any way that identifies you as an individual. Only aggregated data will be presented in the final report. Your participation in this survey is completely voluntary and will not affect your grades. Your time and cooperation are greatly appreciated. If you have any questions regarding this survey or the study in general, please contact: Alysia Johnson, Intern, Baker High School, 123 Main Street, Los Angeles, California, 90013.

Whether you use a cover letter or an introductory statement, the following information must be present.

- *Purpose.* You must provide the participants with information about both the purpose of the survey and the purpose of the research project as a whole. If your survey is asking participants about a prior experience, such as a professional development training or a new intervention strategy program, you must clearly indicate that the purpose of the study is to collect information and/or opinions related to the specific prior experiences. As the researcher, you must help to place the participants in the proper context for answering the questions. Don't assume that participants will make connections to those specific past experiences without being instructed to do so. Such assumptions may result in invalid data that does not reflect the participants' true beliefs or understandings about the experience or phenomena you are interested in studying.

- *Confidentiality statement*. As you likely remember from our ethics discussion in Chapter One, an important duty of a researcher is maintaining **confidentiality**. In order to do this, the researcher must ensure that what a participant says or reports will not be shared with anyone in any way (this also includes the write-up of the study) that would result in an individual being identified. Although this sounds simple enough, survey researchers are often faced with the dilemma of not being able to report certain findings because individuals could be identified and what they said directly attributed to them. This can be a considerable problem when surveying in a small school district or other setting where there may be only one or two individuals in a certain position. Take for example a one-building K–12 school district in a rural community (and yes, they do exist). In these districts, typically there is only one teacher per grade level. What if the findings indicated that the gym teacher, or the English teacher, or the first-grade elementary teacher was disappointed with the administration for its inability to address issues surrounding students with special needs? The researcher would not be able to reveal this specific information for fear of directly violating confidentiality. In larger settings, such concerns pose less of a problem. For example, take a large urban school district with many buildings at the elementary, middle, and secondary levels. A researcher could comfortably report that 50% of the first-grade teachers in the district believed that the administration is not meeting the needs of all students. With such a large number of teachers in the district, it would be virtually impossible to trace this finding back to the specific teachers with such concerns. Researchers should always carefully consider the issues that surround confidentiality. See Box 8.1 for a research example.

Often students confuse confidentiality and anonymity. Unlike confidentiality, *anonymity* means that the survey does not require the participants or **respondents** to provide their names or any information that identifies them in any way (for example, grade level currently teaching, number of years in current teaching position, name of school building, and such). An anonymous survey is one that contains no demographic information that could be used to identify an individual. Making a survey anonymous may increase the number of surveys that the researcher receives back. In addition, participants are more likely to provide valid information if they believe there is no way that their answers and comments can be traced back to them. While these two benefits are good reasons for using anonymous surveys, they make it impossible for a researcher to follow up with participants to collect further in-depth information or validate what participants have said.

BOX 8.1
Ethical Issues in Descriptive Survey Research

Surveys in educational settings are often used to assess the incidence of self-reported problems in youth. For example, the Youth Risk Behavior Survey (YRBS), which is administered in hundreds of high schools each year, asks youth to report on how frequently they use drugs, engage in sexual activities, smoke, feel depressed, or engage in binge eating or vomiting to control their weight. The YRBS was designed by researchers at the University of Michigan to monitor changes over time in these risky behaviors. Surveys are also used in more local studies to determine the extent of problem behaviors in a school or community or to monitor changes in problems when a prevention program is introduced. These surveys have the potential to reveal problems that were previously unknown.

An ethical dilemma that confronts survey researchers is that students revealing problems cannot be helped unless confidentiality is violated. Remember that confidentiality means that student responses cannot be shared with anyone, including parents. If a survey revealed that several students at a school were seriously depressed to the point of thinking about suicide, what might the researcher's ethical responsibility be?

Legally, researchers are not required to report this type of problem. However, there are a variety of ways in which they might address the ethical concerns about doing nothing to help a student who might hurt himself or herself. Some researchers might provide a general report to the school indicating that problems exist in certain areas and recommending that the school promote specific counseling services to address these. In fact, some researchers might work with the school in advance to set up workshops or assemblies to promote these services before the survey is administered. If the students' answers indicated a need for more immediate attention, the researcher would relay this information to the school administrator.

- *Volunteerism*. In the introductory statement or cover letter, the researcher must inform participants that participation in the study is entirely voluntary. Participants must be informed that their lack of participation will not result in negative consequences. Alysia made this clear to the students in her study. She told them that failing to participate would not influence their grades. In survey studies in which the participant has no real "buy-in" (a stake in the issues being explored by the survey), the decision to participate or not in the study would likely have little if any consequence. However, if a researcher were surveying people receiving social services, or mental health services, respondents might fear that participation or nonparticipation could result in loss of their services. In this type of situation, a researcher must include a sentence that clearly states that participation or lack of participation will not affect receipt of services. Additionally, the researcher must inform participants that they have the right at any time to withdraw from the study or may choose not to answer any question.

- *Contact information.* The cover letter or introductory statement must provide contact information for use by participants in the event they have a question or need further clarification on an item. The name of the **principal investigator** (also called the P.I.) should be listed somewhere, along with the name of the institution, group, or agency sponsoring the study (if there is one), a physical mailing address, phone number, and e-mail address (if available).

The cover letter or introductory statement must also include instructions on how to return the survey (for example, there is a self-addressed stamped envelope enclosed) and a deadline date for return of the survey (on average, 2 weeks following the receipt of a survey is adequate; anything shorter or longer can lower response rates). The researcher can include additional information specific to the study that he or she believes to be important and, finally, should thank the participants for their time and cooperation.

Demographics The next section of a survey is demographics. **Demographics** are descriptors that provide detailed information about participants in the study. The specific demographics you gather depend upon the study you are conducting. If your participants are teachers, you might want to know what grade levels they teach, how long they have been teaching, their area of expertise, and so on. These are the typical demographics of interest to educational researchers. Exhibit 8.2 presents an example of demographics that Alysia might gather for her study on students and the format that might be used in her survey.

EXHIBIT 8.2
Sample Demographics Form

About You:

Gender: ____male ____female

Age: __12–13 ____14–15 ____16–17 ____17–18 ____other

Grade Level Taught: ___9th __10th ____11th ____12th

Years in Public School (Fill in): _____Yrs.

Have you ever worn a school uniform? ____yes ____no

When deciding on the type of demographics you will include in your study, consider the following guidelines:

- *Demographics should be derived from the literature review.* When reviewing past studies on your topic, be sure to note what demographics have been used in the past, as well as demographics that "emerge" from the literature. For example, you are interested in the topic technology integration and teacher **pedagogy** (beliefs and practices involved in teaching). In an examination of the literature, you discover that there is a general consensus in the education field that the use of technology by teachers is somehow related to the length of time one has been a teacher. Teachers who have been teaching more years are less likely to integrate technology into their teaching than new teachers. In creating demographics for the survey, you would make sure that you have items gathering information on the *number of years one has been teaching*. Otherwise, you will have hundreds of surveys from teachers but no way to determine whether there was a difference between those teachers who have been teaching longer and those new to the field.
- *Demographics should be used for specific purposes and not as a "shot gun" approach to gathering all possible information on participants*. Many researchers pack as many demographics into a survey as they possibly can. Caution should be taken when selecting demographics to include. Too many demographic items may frustrate participants and negatively affect their desire to fill out the survey. Some participants may have concerns about how demographic information will be used. For example some participants might refuse to give information on race and ethnicity because of concerns that it might be used in a way that makes their group look bad.
- *Demographics can go at the beginning or end of the survey, depending on their purpose.* Although demographics traditionally appear at the beginning of a survey, in some situations it may be more appropriate to end the survey with demographic items. The decision for putting the demographics later rather than sooner depends on how important these demographics are to the researcher and to the study. If demographics are imperative to the study, then they should go at the beginning of the survey, when the participants are fresh and ready. If demographics are not imperative, they should go last, allowing the participants to exert most of their energies (and brain cells) on answering the survey items.

The Body of the Survey The development of individual survey items is often confusing to those who have not created a survey. Students often mistakenly think that the subquestions *are* the survey questions. As we noted earlier in the chapter, the subquestions are written at the end of the literature review and represent the main sections of the survey. These sections are developed using the subquestions. The researcher then uses the subquestions to generate the survey items.

For example, in Alysia's study on student perceptions of school uniforms, one of her subquestions involves how students believed school uniforms would influence their school community. As in the example provided in Exhibit 8.3, the section of school community is one of the three sections of the survey. Notice how all the items in that section pertain to school community and solicit a *quantitative* response.

The number of individual survey items in each section depends on the topic and can vary. There is no "definitive" number of items that make up a survey. In general, for the purposes of the research proposal in this course you should aim for 8–12 items for each section.

Criteria for Writing Good Survey Items As discussed earlier, descriptive survey researchers usually develop their own instruments. Sometimes they may borrow or modify items from surveys developed by other researchers. One should always ask for permission to use or adapt items from other researchers. Regardless of whether items are developed by the researcher or borrowed, all items should meet certain criteria for good survey items.

- *Be clear and concise in the language used*. Define all terms that could be interpreted differently than you intend. For example, if you are interested in studying

EXHIBIT 8.3
Student Perceptions of the Impact of School Uniforms on the School Community

Read each item below and respond by circling the number that represents how you feel.

1 = Strongly Disagree; 2 = Disagree; 3 = Slightly Disagree; 4 = Slightly Agree; 5 = Agree; 6 = Strongly Agree

Because of school uniforms . . .

I think we will see positive changes in the behavior of students in the school cafeteria.	1 2 3 4 5 6
I believe there will be less competition among students.	1 2 3 4 5 6
I think that students will get along better.	1 2 3 4 5 6
I believe my school will be a better place.	1 2 3 4 5 6

teachers' use of *constructivist* instructional approaches, it would be better to reword this term or define it. If you are ultimately interested in learning about teachers' use of hands-on instruction, then you should modify the wording so that it is more specific and clear, and use the words "hands-on learning."

- *Make sure each survey item gathers data on one central idea or question.* Although it is tempting to ask participants two questions in one item to save time and space, this will result in confusion, and it will be impossible to interpret responses. For example, consider an item asking participants to respond on a scale of "strongly agree" to "strongly disagree" to the following statement: *I believe that school uniforms will make the cafeteria and the halls more friendly.* If a participant agrees strongly with this statement with regard to the cafeteria, but disagrees with regard to the halls, how should the participant respond to this item? Strongly agree? Strongly disagree? Not Applicable? The same problem arises in the second part of the question. Cafeteria or halls or both? To fix these problems, the researcher must break this question down into two separate questions, each containing one essential idea.
- *Avoid use of double negatives.* For example, *It is not good practice for teachers to not meet the needs of all their students.* As you read this item, you probably felt yourself stop and think about it for a second before answering it. Right? It is very difficult to determine what a response of Strongly Agree or Strongly Disagree would mean in this case.
- *Make sure items (particularly demographic items) have response sets that do not overlap.* This is a common mistake, even in professional surveys. Consider a survey item asking participants to indicate annual income levels. If you have selections that overlap, as in the example following, it will be virtually impossible for you to decipher who has selected the correct response for their income level.

Indicate your income level by checking one of the categories below.

- ❒ \$0–\$10,000
- ❒ \$10,000–\$30,000
- ❒ \$30,000–\$50,000
- ❒ \$50,000–\$80,000
- ❒ \$80,000–\$100,000
- ❒ \$100,000 or more.

If you were filling this demographic item out and you earn \$50,000 annually, which one would you select, \$30,000–\$50,000 or \$50,000–\$80,000? What if, like you, 99% of the sample earned \$50,000, and the rest of the sample selected the first choice but you selected the second? The finding that 99% of the sample earned $30,000–$50,000 and only 1% earned $50,000–$80,000 would be incorrect. In fact, in this example, the entire sample earned $50,000 that year.

- *Include all possible responses*. For example, if you are asking participants how many times that week they used technology in their classrooms and you provide a frequency checklist like the one following, it is important that you provide a zero or none category.

How many times did you integrate technology into your lessons last week?

- ❐ 0
- ❐ 1–3
- ❐ 4–6
- ❐ 7–9
- ❐ 10–12
- ❐ 13 or more.

If you did not provide a "zero" as a possible answer, many of your participants may not be able to provide an accurate answer (assuming someone never used technology), and the data will overestimate the degree of technology integration. If the zero response is not provided, and 100% of teachers filling out the survey had not used technology in their teaching the week before, they might either leave the question blank or might select the item that most closely represents their situation (1–3 times). If 60% chose 1–3 times, the researcher would report that 60% of respondents used technology 1–3 times in the week prior to being surveyed, when really none had.

"All possible responses" also includes use of the category "Other." The benefit of a researcher providing the category Other is that it opens the response so that the participant can provide the most accurate response possible. If you use the category Other, be sure to provide a place for the participant to write in a detailed response. However, if an overwhelming number of respondents select the Other category on many different items, this could mean that the researcher has not properly identified the sample of participants or has not included viable responses for all participants. Remember, it is from a thorough review of literature that you will determine the viable responses to your questions.

- *Write items that do not assume information about the participants*. In today's very diverse society, it is easy for researchers to make assumptions about the backgrounds of participants. For example, questions that are related to family should not assume that every family has two parents, working parents, siblings, and so forth.
- *Write items that allow participants to express what they believe rather than suggesting a particular answer*. Researchers may inadvertently use a survey to obtain data to support

their preexisting beliefs, writing items in such a way as to suggest that certain responses are "better" than others are. Items that are written in a way that lead participants to the response that the researcher wants them to give (even though the researcher is not aware of this) are referred to as **leading questions**. For example:

Educational research, one of the most important classes in graduate school, should be how many credits?

- ❒ 1
- ❒ 2
- ❒ 3
- ❒ 4

This is a leading question because you are making an assumption that educational research is one of the most important classes in graduate school. (But isn't this true for you?)

Step 5: Select the Participants

Ideally, the survey researcher wants to send surveys to every member or individual within the chosen populations. Although these large populations are referred to as **ideal populations**, sampling every person in these populations is not realistic or "doable." Time, money, and other resources such as staffing typically make it impossible for the average survey researcher to reach all members of the ideal population. Therefore, the researcher has to forgo these grand expectations and select a smaller or **realistic population**. Presented in Box 8.2 are samples of ideal populations, followed by realistic populations.

BOX 8.2
Samples of Ideal and Realistic Populations

Ideal Population: Fifth-grade teachers in the United States

Realistic Population: Fifth-grade teachers in New York City

Ideal Population: High school principals on the east coast

Realistic Population: High school principals in Pennsylvania

Ideal Population: Parents who have children in day care in Chicago

Realistic Population: Parents who have children in day care in three neighborhoods on the west side of Chicago

In some situations, even these "realistic" populations may not be so realistic. For example, surveying all the fifth-grade teachers in New York City or high school principals in the state of Pennsylvania might exceed the study's resources. In addition, the researcher must be able to obtain a complete list of persons in the realistic population. Some researchers refer to this as the **sampling frame**. For teachers, administrators, and school districts, researchers can consult their state education department and state educational organizations to obtain these lists. For other populations, local, state and national organizations may help to identify possible participants. For some studies, such lists do not exist. For example, a college counseling center might want to identify participants for a study of students who are depressed but have not previously sought services. In this case, the center might run an advertisement in the student newspaper asking for participants for their study.

After you have identified a list of possible participants, the next step is to select a **sample**. A *sample* is a smaller group selected from a larger population (in this case, a realistic population). Samples allow researchers to work with a smaller, more manageable group out of the realistic population. There are two major ways that researchers select samples, random and nonrandom, which are discussed following.

Types of Random Sampling The most important aspect of sampling is that the sample must represent the larger population from which it is drawn. **Random sampling** is a technique or tool that produces essentially a mini-version of the initial population. Random sampling is conducted in such a way that every person in the population has an equal and independent chance of being selected. This means that when a person is selected, it does not affect the chances of anyone else being selected. Take, for example, a technique often used in grade school spelling lessons. Remember your elementary teacher asking everyone in your class to form a line and to count off by ones and twos to establish teams of spellers? Is this an example of random sampling? No, the moment you got in line with your classmates and the teacher said "Okay, count off by ones and twos," your group membership was decided. If your best friend, immediately to your left, was a one, then you had to be a two—and there was no chance that the two of you would be on the same spelling team. This scenario violates the requirement that selection of one person cannot have an effect on whether another person is chosen or not.

- *Simple random sampling.* **Simple random sampling** involves the random selection of individuals from the realistic population as a whole. First, the researcher obtains a complete list of names for all those individuals that make up the realistic population. To select a simple random sample, each person on the list of the realistic population is assigned a number. For example, if the list contains the names of 20 individuals, then the number 01 is assigned to the first person on the list, and number 02 to the second person on the list, and so on, until all 20 people on the list have been assigned a number. Next, a random sampling table (usually generated by computer) is used. Random number tables present clusters of number strings that have been randomly generated. In Table 8.2, a list of random numbers from a random sampling table is presented in the third column.

To use the list, begin anywhere you want and select a number. If number 05 is selected, person 05, Derrick, is selected for your sample. Next you can move up or down or in any direction on the list to select your next number, because, after all, the list is random. For example, if you move down the list of random numbers,

TABLE 8.2 Simple Random Selection

Possible Participants		List of Random Numbers
01	Steve	22
02	Tonya	05
03	Ramone	18
04	Juan	04
05	Derrick	07
06	Tony	18
07	Liz	16
08	Jean	14
09	Dean	48
10	Margie	41
11	Kate	17
12	Maria	12
13	Ismael	12
14	Jim	10
15	Donna	19
16	Heta	45
17	Avi	27
18	Richard	32
19	Ron	08
20	Travis	18

the next number is 18, so Richard would be added to the sample. If a number is selected that is not on your list of individuals in the population (such as 48 in our example), simply move to the next line of the random numbers. Continue this process until you have your entire sample.

- *Stratified random sampling.* There may be times that a simple random selection will not generate the type of participants needed in a sample. For example, if a researcher conducting exit polling during an election uses a simple random sampling procedure to choose individuals to survey as they leave the polls, just by chance he or she would not get a representative sample of Democrats, Republicans, and Independents. For such a poll, a researcher would want to obtain a sample that was proportional to the number of Democrats, Republicans, and Independents in the geographic area being studied. A **stratified random selection procedure** allows the researcher to stratify along the variable of party affiliation to select a sample that is more representative of the population. If possible, the researcher would obtain a list of eligible voters according to political party and then randomly select the appropriate proportional number of participants from each group. Whenever subgroups are critical to creating a sample that represents the entire population, stratified random sampling is the most precise sampling technique. Variables that are used to stratify a sample in educational research might include race, ethnicity, socioeconomic status, years of teaching experience, grade level or school location: urban, suburban, and rural.
- *Cluster random selection.* In the field of education, simple random selection is often not possible. For example, if you are surveying teachers in a particular state, you might not be able to obtain a list of individual teacher names but you can get a list of school buildings for the state. In this case, **cluster random sampling** may be useful. Instead of assigning numbers to individuals, in cluster random selection, numbers are assigned to the cluster or "subgroup" within the realistic population. In the previous example, teachers in each school building would be considered a subgroup (or cluster). The building would be assigned a number, and buildings would be randomly selected using the random number table process described previously. Then all persons in the cluster would be included in the sample. By selecting clusters, you reduce the number of schools you need to visit or the number of contact persons you need to identify. The important thing to remember is that cluster random selection is a procedure through which entire groups and not individuals are randomly selected. This procedure allows the researcher to select clusters randomly and is a simpler technique than selecting individuals randomly. However, to be certain that the sample accurately reflects the population, a researcher would likely have to select multiple clusters. The number of clusters you select would be determined by the number of clusters in the population.

Random selection of the sample allows a researcher to generalize the results of the study back to the entire population from which the sample was drawn. Sometimes researchers use a **nonrandom sampling** procedure. Because of limited time, resources, and/or purpose, a researcher may conduct a study in which teachers or students in only nonrandomly selected school buildings or districts are included in the study. This type of nonrandom sampling is referred to as convenience sampling. Although a sample of convenience requires fewer resources, it severely limits a study's generalizability. In this case, the study's results are only "good for" or generalizable back to the teachers in the school building or district. Depending on the overall purpose of the study and how the results of the study will be utilized and disseminated, lack of generalizability might not be an issue. However, if the intent of the study is to add to the academic or professional knowledge base on a subject by publishing on a national level, such sampling techniques would be considered a major flaw in the study design.

Census sampling is another nonrandom sampling technique used in quantitative research. In census sampling, the researcher surveys the entire realistic population without drawing a random sample from the population. This technique may be used when the study has unlimited resources and/or the realistic population is not too large. Census sampling is frequently used by educators who are trying to obtain data only on their own school or district. Such data can be useful in learning about that school or district; however, remember that the results cannot be generalized to other schools or districts because the sample was not chosen randomly.

While random selection plays a pivotal role in a study's credibility, the size of the sample that one selects from the realistic population is also very important. If the sample is too small, it may not fully represent the population from which it was drawn, and the findings from the study cannot be generalized back to the wider audience even though random sampling practices were utilized.

Even though there are no "hard or fast" rules for determining sample sizes, there are some general guidelines to consider when planning a study. For survey research, if the population is fewer than 200 individuals, the entire population should be sampled. This would be considered census sampling. At around a population of 500, approximately 217 participants from the population should make up the sample, and populations over 1,000 require approximately 278 for an appropriate sample. For very large populations of 5,000 or more, samples of 350 to 500 persons are often adequate (Dillman & Salant, 1994).

Step 6: Pilot and Administer the Survey

Any survey should be pilot tested with a small group of persons similar to those who will be in the final sample. Think of piloting as a kind of "dress rehearsal" for a survey. Participants that make up a pilot sample are usually chosen at random and are given the survey to complete, and they are asked to examine the survey on many different fronts: clarity of language and terms, spelling and grammar, depth and breadth of sections and the items that make up those sections, and overall psychometric properties of the instrument (for example, scales are correct, and so on). One hopes that the survey will be sound enough for members of the pilot group to complete it. When piloting your study, provide an additional sheet to the survey on which pilot participants may write any comments, suggestions, or questions they have about the survey. Like Alysia, you should use this feedback to make corrections or refinements to the final survey.

After piloting the survey the researcher must decide how to administer the survey. The paper-and-pencil, mail-out, mail-back survey is the traditional method for survey administration. This method has many benefits, especially from a measurement perspective. First, it helps to ensure that the confidentiality of the participants' responses is maintained. The researcher mails surveys directly to a group of participants from the realistic population, who fill them out and mail them back using an enclosed self-addressed, stamped envelope. This helps to ensure that no one except for the researcher has access to the information on the survey.

This method of administration also has disadvantages. With large samples, the cost of postage may be a factor. Also, in order to use this method, the researcher must have mailing addresses for participants. For example, when surveying teachers, obtaining home mailing addresses is unlikely. However, it may be possible to send surveys to each teacher at a school building.

As discussed earlier, using the telephone to assist in the delivery of a survey is an alternative method for survey administration. The benefit of this approach is that the researcher can collect data within a relatively short period of time than with the mailing method, which may take 3–4 weeks for responses to come back. Phone surveys allow the survey administrator to encourage participation and increase response rates by personally explaining the purpose of the survey, answering questions about the study, and establishing rapport. A problem with using the telephone, however, is that the researcher must have the phone numbers of the participants. Unless participant phone numbers are readily available, researchers should avoid this method. Looking up phone numbers, unless they are grouped somehow in a directory, is too time consuming and not worth the effort. Marketing firms sometimes generate phone number samples randomly, but this approach is not likely to produce a sample representative of any meaningful educational population. Timing is another problem with using the telephone for educational research surveys. If contacting teachers, do not expect to reach them during the day at school, even though most schools have phones in classrooms and teachers have voicemail. Leaving messages either on voicemail or with the main office typically results in low response numbers. In this situation, it would be better to try to contact the teachers at home.

Although other, less "controlled" approaches are not ideal, sometimes researchers use them to administer and collect surveys. If, for example, a researcher is surveying teachers in a school district or building and doesn't know the names of the teachers (or how many teachers are in the school), the researcher can identify a key person within the school to assist in coordinating such efforts. In this case, the researcher sends a package of surveys to the key contact person at the school, and this "insider" coordinates the efforts, giving the surveys to the individual teachers. Once the teachers have filled out the surveys, the contact person collects and mails them back to the researcher. With this method, the researcher spends very little effort in the administration of the survey; however, the **validity** of the data can be severely compromised, depending on the selection of the key contact person and that person's relationship to the participants. Imagine that a researcher has selected the school principal as the key contact person. Now, try imagining that you are a teacher dissatisfied with working at the school. Would you be willing to give your boss a survey in which you have indicated your dissatisfaction? Having the principal give out the surveys, then collecting completed surveys in a secure drop box within the school building might provide a solution to such a problem.

An increasingly popular method of obtaining survey data is through the use of the Internet. Surveys can be e-mailed to groups and responses returned electronically, or a group can be invited to visit a Web site to complete a survey. One advantage of online surveys is that the survey can be set up so that data is automatically entered and tabulated.

Internet surveys are most valid and effective when the researcher can obtain a list of e-mail addresses for the realistic population. Another approach is to send a message to a listserv or electronic discussion group that focuses on the issues addressed by the survey.

Step 7: Analyze and Interpret the Survey Results

Descriptive survey research often produces huge amounts of data, in part, because large samples are used. A major goal in data analysis is to organize and summarize the data using descriptive statistics.

Let's illustrate what data from a descriptive survey study might look like using 10 students from the 1250 respondents in Alysia's study. After Alysia collects her surveys, she sets up a database to conduct an analysis. She records participant responses for each question. As you can see, there is variability in her data. Student #1 strongly disagreed with item 1 while Student #6 strongly agreed. For her project Alysia decides to run some basic descriptive statistics: mean, standard deviation, and percentages. She calculates the percentage of students who agree with each item (rating 4, 5, or 6) and the percentage of students who disagree with each item (rating 1, 2, or 3). Presented in Exhibit 8.4 are four items from

EXHIBIT 8.4
Database for 10 Students from Alysia's Study on School Uniforms

Item 1: I think we will see positive changes in the behavior of students in the school cafeteria.
Item 2: I believe there will be less competition among students.
Item 3: I feel that students will get along better.
Item 4: I believe my school will be a better place.

1=Strongly Disagree; 2=Disagree; 3=Slightly Disagree; 4=Slightly Agree; 5=Agree; 6=Strongly Agree

Responses of 10 students to Alysia's survey

Student	Item 1	Item 2	Item 3	Item 4
#1	1	1	1	1
#2	5	5	4	2
#3	2	2	2	5
#4	1	4	3	3
#5	1	1	1	1
#6	6	6	6	5
#7	5	4	5	4
#8	1	3	5	6
#9	6	6	6	6
#10	1	1	1	1
Mean:				
% of Students in Agreement				
% of Students in Disagreement				

Alysia's survey, the 6-point Likert scale used by students, and a database containing the responses from 10 of the 1250 students. Conduct a basic analysis of this data for yourself.

Once the survey has been administered, the final concern of the survey researcher is how to interpret the results. The researcher would ultimately want to generalize the findings of the study back to the realistic and ideal populations. The researcher must consider the number of responses received or the **response rate**. *Response rate* is the percentage of persons in the sample who complete and send surveys back to the researcher. A researcher never gets back 100% of the surveys sent out. Response rates vary considerably depending on the purpose of the study, the relationship between the participants and the researcher, and the subject of the survey. Even after several reminders, response rates of 30% to 50% are typical. Response rates are higher in situations in which the people being surveyed have a greater interest or "stake" in the topic or results, for example, perhaps they attended a 3-week symposium on the topic the summer before, or the survey results may lead to changes in the curriculum that they are teaching. In these situations, response rates may be 80% or higher.

Some methods of survey administration, such as Internet surveys, have particularly low response rates. For example, Kennedy and Voegtle (2000) surveyed ally groups at junior and senior high schools by posting their cover letter on the listserv of an organization that had set up more than 600 ally groups for gays and lesbians. The cover letter invited members of the listserv to participate in their study and provided a password for accessing the survey (an important step for controlling the sample). Kennedy and Voegtle (2000) had a response rate of less than 1% despite repeated reminders to participants to complete the study survey. (Since this was not likely to be a representative sample, these researchers switched to using a qualitative approach. See chapter 5 for a discussion of qualitative research methods.)

Low response rates can have serious implications for the generalizability of the results of a survey study. Even when the sample to which the survey was distributed was randomly selected, a 50% response rate may result in a final sample that is not representative of the population. Often people who have extreme opinions about an issue (either in favor of or against it) take the time to respond to a survey, and those in the middle (not really caring that much one way or another) may be less likely to take the time to fill the survey out and return it. If a majority of the randomly selected sample for a survey does not respond, the findings reported may be misleading; if everyone who had been part of the sample had responded, the findings might be quite different. In general, professional journals consider for publication only studies for which the response rate is 50% or higher and preferably 70% or higher.

EVALUATING DESCRIPTIVE SURVEY RESEARCH

While every study has its own purpose, there are some basic elements involved in evaluating survey research. Following are some criteria often used to evaluate descriptive survey research.

Criteria for Evaluating Descriptive Survey Research

Presented in Table 8.3 are some key criteria to use when you are evaluating the rigor of a descriptive survey study.

TABLE 8.3 Criteria for Evaluating Descriptive Survey Research

Criteria	Details
Random selection from a population	A rigorous survey study will use a type of random selection to select its sample from the larger population.
Pilot test of survey	It is important that the study pilot the survey with a small group, drawn either randomly or through convenience, to pilot or try out the survey before administering it to the entire sample.
Modifications to survey	As a result of the piloting process, the researcher should discuss changes made to the original survey.
Establishing validity for the survey	The researcher should discuss how content validity and face validity were established for the survey. This may be done simultaneously with the pilot sample and/or through a panel of experts examining the survey items.
Method for survey administration	The researcher should discuss how the survey is administered to the participants (for example, through the mail, e-mail, handed out in person).
Response rate	The researcher should also disclose the response rate or the number of people who returned the survey. Response rates of less than 50% are subject to criticism. Most journals require 70% return rate or greater.
Method used for improving response rate	To increase low response rates, the researcher should conduct a second mailing or other methods to improve response.

SUMMARY

Descriptive survey research is one of the most common types of quantitative research in education. There are different types or designs of survey studies. The most common design is the one-shot design, in which the researcher, after piloting the survey, administers it once to the sample. Follow-up studies and longitudinal survey studies are designs that follow the same sample over time. Designs that select different samples over time include trend and cohort surveys. Designs that follow the same or different samples over time both require surveys to be administered multiple times.

Researchers use cover letters to accompany their surveys. The cover letter defines the purpose of the study, discusses confidentiality of data, and provides the researcher's contact information. The survey itself is composed of different sections that gather different types of data. The first section of the survey is called the demographic section and collects personal information about the participants. The body of the survey is made up of similar items grouped together. These groupings are aligned with the research subquestions the researcher is asking. When writing survey items, the researcher should adhere to some basic criteria. Survey items should be written in clear, concise language. Items should gather data on one central idea or question and should avoid leading statements and containing double negatives.

In selecting the sample for a survey study, researchers often use census sampling or select random samples out of the realistic population in order to have the widest generalizability possible. Before surveys are administered to the sample, a pilot study is typically conducted. The pilot study selects a smaller number of individuals from the sample and conducts a test run with the survey to "get all the kinks out." The researcher will use feedback from participants in the pilot study to make modifications to the survey before administering it to the remaining sample. Surveys can be administered to participants using several methods. While alternative methods for survey administration such as e-mail, phone, or in-person are used, the traditional mail-out, mail-back method continues to be the most widely used procedure.

Survey researchers need to be concerned about the response rate or number of surveys that participants fill out and return. A study that has less than a 50% response rate may be reporting findings that are skewed and do not accurately reflect the beliefs or attitudes of the total sample who received the survey.

KEY CONCEPTS

census sampling
cluster random sampling
cohort survey design
confidentiality
demographics
descriptive comparative
follow-up study
ideal population
leading questions

longitudinal survey study
nonrandom sampling
one-shot survey design
pedagogy
principal investigator
questionnaire
random sampling
realistic population
response rate
respondents
sample
sampling frame
simple random sampling
stratified random selection procedure
subquestions
survey
trend study
validity

DISCUSSION QUESTIONS OR ACTIVITIES

1. A researcher is conducting a study that examines the perceptions of patients receiving mental health services. The purpose of this study is to describe the types of services these patients are receiving and whether they believe these services are affecting their quality of life. The researcher is planning to administer the surveys with the help of the patients' counselors. Discuss any benefits or limitations that you see in this researcher's approach, and what, if any, alternative methods you could suggest to help improve the study's methodology.
2. A researcher is interested in surveying fourth graders for the purposes of describing how they feel about high-stakes testing, and how, if at all, they prepare for these tests. From a developmental perspective, discuss some challenges this researcher might encounter in designing such a survey for students at this age.
3. A researcher is surveying teachers in an urban school district as to whether they believe a new elementary math program is effective. The math program is a controversial subject among many of the school staff. In all, the district has six elementary buildings. The researcher has decided to use anonymous surveys. From a measurement perspective, discuss some of the benefits and limitations of this approach in reporting findings.

SAMPLES OF DESCRIPTIVE SURVEY STUDIES

Abrams, L. M., Pedulla, J. J., & Madus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory Into Practice, 42*(1), 18–29.

Brown, M., & Bergen, D. (2002). Play and social interaction of children with disabilities at learning/activity centers in an inclusive preschool. *Journal of Research in Childhood Education, 17*(1), 26–37.

Cameron, L., & Lee, B. (2009). The researchers ate the homework! Perspectives of parents and teachers. *Education Canada, 49*(1), 48–51.

Cho, S., & Reich, R. A. (2008). New immigrants new challenges: High school social studies teachers and English language learner instruction. *Social Studies, 99*(6), 235–242.

D'Andrea, M. (1994).Concerns of Native American youth. *Journal of Multicultural Counseling & Development, 22*(3), 173–181.

Rose, S. E., Jolley, R. P., & Burkitt, E. (2006). A review of children's, teachers' and parents' influences on children's drawing experience. *International Journal of Art & Design Education, 25*(3), 341–349.

SUGGESTED READINGS

Fink, A. (2006). *How to conduct surveys: A step-by-step guide.* Thousand Oaks, CA: Sage.

Groves, R. M., Fowler, F. J. Jr., Couper, M. P, Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology.* San Francisco: Jossey-Bass.

CHAPTER NINE

EXPERIMENTAL RESEARCH

CHAPTER OBJECTIVES

- Identify the distinctive characteristics of experimental research
- Explain the differences among directional, nondirectional, and null hypotheses
- Explain the operational definition of a variable using examples of such definitions for both a dependent and an independent variable
- Define extraneous variables and describe the main techniques and designs used by experimental researchers to control them
- Define seven threats to both internal and external validity and explain ways to control for each threat
- Explain the different types of experimental designs
- Explain methods used by researchers in single-subject designs

RESEARCH VIGNETTE

Johanna is a graduate student in secondary education who is interested in math achievement and computers. Her literature search has shown that there is a relationship between math achievement and the number of computerized tutoring sessions that a student receives. Much of the research she has found suggests that further study should use methods that will allow researchers to conclude that computerized tutoring sessions are related to increased math achievement. Knowing that it's the strongest way to demonstrate a cause-and-effect relationship, Johanna decides to conduct an experimental study. She realizes the difficulty of conducting research in schools; her part-time professor is the superintendent of a local school district, and she hopes he can help. Her professor, after reviewing her institutional review board proposal, agrees to allow her to conduct the experiment. Johanna randomly selects 60 seventh graders from a pool of 175 students enrolled in an after-school tutoring program. After Johanna randomly selects her 60 seventh graders, she randomly assigns 30 students to the computerized tutoring group and 30 to a noncomputerized tutoring group that receives tutoring from a teacher (the number of computers available is limited such that only 30 students can have access to computers at one time). She selects a teacher identified by the principal as an excellent math teacher to run the noncomputerized group. Students in the computerized tutoring group are supervised by but receive no instruction from the teacher. In the computerized tutoring session, the teacher answers questions related to working with the computer. In the noncomputerized tutoring session, she provides materials with math problems and works individually or in small groups with students. Participants in both groups are pretested using a math achievement test, and then both groups attend 30-minute after-school sessions for 30 weeks. (A *pretest* is a test given before the experimental treatment in order to see if the groups are equal.) The groups are then posttested with an alternate form of the same test. (A *posttest* is a test given after the experimental treatment.) The results indicate that the computerized tutoring group outperformed the noncomputerized tutoring group on the posttest.

UNDERSTANDING EXPERIMENTAL RESEARCH

Experimental research, which comes out of the framework of scientific realism, is thought by many to be the only type of research that results in findings that suggest causal relationships. What makes experimental research distinct from other forms of quantitative research is that the researcher controls or manipulates how groups of participants are treated and then measures how the treatment affects each group. In technical terms, the researcher controls or manipulates one or more independent variables and examines the effect that the experimental manipulation has on the dependent variable or the outcome of the study.

Remember from our discussion in chapter 2 that the **independent variable** is the variable that refers to how participants are treated. Participants are usually assigned to different groups that receive different treatments. In the field of education, the independent variable might be curriculum materials (e.g., skill-based readers versus literature), instructional styles (e.g., group learning versus individual), or specialized training (e.g., receiving training or not), to name just a few. The outcome of the study is the **dependent variable**, which is typically measured by a test or a measuring instrument that produces quantitative data. Table 9.1 shows the independent and dependent variables for Johanna's study.

Consider as an example Johanna's study on math achievement and a computerized tutoring program. In such a study, the independent variable would be the type of tutoring. The researcher would decide (or manipulate) the individuals who would receive the computerized tutoring and those who would receive tutoring from a teacher. In other words, the researcher would expose one group of participants (the experimental or treatment group) to computerized tutoring and the other group (the control or comparison group) to noncomputerized teacher-led tutoring. Math achievement, the study outcome, would be the dependent variable and would be used to determine whether any difference resulted from the manipulation of the independent variable (type of tutoring).

Simultaneous to the consideration of the independent and dependent variables, researchers must consider any potential variables that could influence the groups' performance on the dependent variable. Remember that at the end of a study, the researcher wants to conclude that the treatment *caused* any differences found between the experimental and control groups. To legitimately do this, the researcher has to be certain that no other variables could cause the differences. Technically, the researcher wants to control for **extraneous variables**, which, as you remember from chapter 2, is a variable that could influence the participants

TABLE 9.1 Experimental Research

Group Assignment	Independent Variable	Dependent Variable
Experimental or treatment group	Computer-assisted tutoring sessions	Math achievement as measured by math test
Control or comparison group	Noncomputerized tutoring sessions	Math achievement as measured by math test

in the study and ultimately influence the dependent variable. Returning to our example of seventh graders and math achievement, if a difference were found at the end of the study, we would want to conclude that the reason or cause of the difference was computerized tutoring. For us to have confidence in this conclusion, our two groups must be as similar as possible, with the exception of the treatment—in this case, computerized tutoring. Johanna examined whether her two groups were equal prior to giving the treatment by pretesting her participants. Ways to control for extraneous variables are discussed later in this chapter.

STEPS IN PLANNING AND CONDUCTING EXPERIMENTAL RESEARCH

Experimental studies closely follow a prescribed set of procedures that are detailed in the research proposal. Once an experimental study begins, there is little deviation from these procedures. The researcher takes an active role in setting up the study but, unlike a qualitative researcher, does not play an interactive role with the participants. As we noted earlier, an experimental study is useful for determining cause-and-effect relationships. For example, the following research questions might lead to studies attempting to establish cause-and-effect relationships:

- Does computer use cause an increase in math achievement?
- Does social skill training have an effect on the communication skills of preschoolers?

Either one of these studies would use the following steps for conducting experimental research:

1. Select a topic.
2. Review the relevant literature and define a research question.
3. Develop a research hypothesis.
4. Select and assign participants to groups.
5. Select measurement instruments.
6. Select controls for extraneous variables.
7. Define and administer the experimental treatments.
8. Collect and analyze data.
9. Make a decision about the hypothesis.
10. Formulate conclusions.

Remember that in preparing a proposal, you will proceed through all but the last three of these steps. In a proposal, these final steps may be replaced by the benefits-and-limitations reflections section of the proposal (see chapter 15). Note, however, that in the method section of the proposal, you will discuss how you plan to collect and analyze the data.

Step 1: Select a Topic

As with other research methods, the experimental researcher bases the topic selection on personal interest, experience, and an initial review of the literature. Typically the researcher is interested in determining whether some treatment causes a significant change in behavior.

Step 2: Review the Relevant Literature and Define a Research Question

The researcher does an exhaustive literature review to determine the findings of current research on the topic of interest. The researcher examines past literature to determine how others have researched the same topic, what variables or issues were studied, discussed, or both, and what the findings of those studies indicated. With this knowledge, the researcher generates a research question and designs the procedures to be used in the study, often borrowing methods used in prior research for his or her own study. An exhaustive literature review is necessary so that the researcher can make an educated and informed prediction (a research hypothesis) about the expected outcome of the study.

Step 3: Develop a Research Hypothesis

Based on the findings of the literature review, the researcher develops and states a hypothesis that indicates the expected causal relationship between the variables. For example, in our study on type of tutoring sessions and math achievement, a **research hypothesis** might be as follows:

> *It is hypothesized that students who receive computerized tutoring sessions will demonstrate a higher level of math achievement than the group receiving noncomputerized tutoring instruction.*

The variables—math achievement and type of tutoring sessions—must then each be operationally defined. With an **operational definition**, a variable is defined in terms of how that variable will be measured, manipulated, or both. In this case, the researcher would define *math achievement* (the dependent variable) as a score on the math test selected to measure achievement (here you would name the test). Independent variables are operationally defined by explaining the procedures used to deliver the different treatments to the experimental and control groups. In our example, *computerized tutoring sessions* might be defined as using computer programs (identifying the specific software) to review math concepts for students during the after-school tutoring sessions. *Noncomputerized tutoring sessions* would be defined as having the teacher present the same math concepts and problems using noncomputerized methods, such as worksheets, during the after-school tutoring sessions.

There are three types of hypotheses used in experimental research, which are described in detail following.

Directional Hypothesis A **directional hypothesis** states the direction or the expected outcome. That is, the researcher feels confident enough to suggest which group would outperform the other group, as in the research vignette at the beginning of the chapter:

It is expected that the computerized tutoring group will perform significantly better than the group receiving noncomputerized tutoring instruction.

Researchers state a directional hypothesis if the literature suggests that there is sufficient evidence to predict the direction of the difference between the groups.

Nondirectional Hypothesis A **nondirectional hypothesis** simply states that there will be some difference between the variables, but the direction of that difference is not being predicted. Let us say that as you review the literature on computer-assisted instruction and math achievement, you find that the evidence is not strong enough to suggest that the computer group will outperform the noncomputerized classroom instruction group. In fact, it is possible that teacher-led tutoring might lead to higher scores. In that case, you might suggest the following hypothesis:

> *It is hypothesized that there will be a significant difference between the computerized tutoring group and the noncomputerized group in math achievement.*

Null Hypothesis. The **null hypothesis** states that no **significant difference** between the variables is expected after the treatment is applied. The null hypothesis is implicit in all experimental research. That is, inferential statistics (discussed in greater detail in chapter 11) always test the null hypothesis, and in most cases, the researcher hopes to disprove the null hypothesis in favor of the research hypothesis. In our example, the null hypothesis would be

It is hypothesized that there will be no difference in math achievement between the computer using tutoring group and the non-computer-using group.

The three types of hypotheses are summarized in Table 9.2.

Step 4: Select and Assign Participants to Groups

Most experimental studies have at least two groups, often referred to as the *experimental* and *control* groups. In an experimental study, the researcher randomly selects and randomly assigns participants to groups. Please note that these are two different processes, as depicted in Figure 9.1.

Remember from our earlier discussion of sampling that random selection allows the researcher to take the findings based on the sample and generalize those findings back to the entire population. For example, let us say that you are interested in conducting a research study on fourth graders in a particular inner-city school district. A list of all the fourth graders in the district would be your defined population. From that population, you would randomly select the desired number of participants to obtain the sample for your study. Although the study is conducted on the sample of fourth graders, ultimately, you want to make statements about the population. The type of randomization procedure used can vary from study to study, but the ultimate goal is still generalization.

For practical reasons, many experimental researchers do not randomly select subjects. The lack of random selection limits the generalizability of results and the statistical tests that can be used. However, random assignment of individuals or groups to treatments is an important feature of experimental research.

Step 5: Select Measurement Instruments

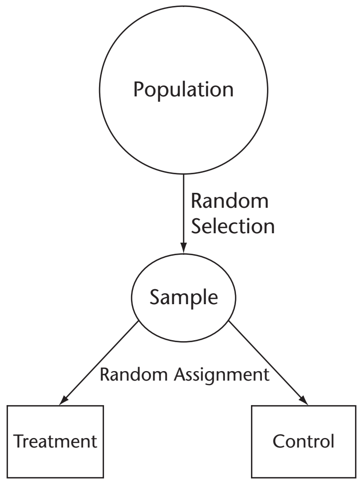
Instruments or measurement tools for an experimental study are selected with the same care and attention as in other types of research. First, you want to be certain

TABLE 9.2 Hypotheses in Experimental Research

Type of Hypothesis	Definition
Directional[a]	States that a difference between the variables is expected and predicts the direction of that difference
Nondirectional	States that a difference between the variables is expected but does not predict the direction of the difference
Null	States that no difference between the variables is expected and any difference found was due to chance.

[a]Often referred to as the research hypothesis.

FIGURE 9.1 Random Selection and Random Assignment



that the instrument you select is an appropriate measure for your dependent variable. Let us say that in Johanna's study, her research hypothesis is

> *It is hypothesized that the seventh-grade students who receive computerized tutoring will have higher math achievement scores than seventh graders who received noncomputerized teacher-led tutoring.*

The dependent variable in this study is math achievement. An appropriate instrument would be a math test that would be valid and reliable for seventh-grade students receiving either method of tutoring.

Step 6: Select Controls for Extraneous Variables

Several techniques for controlling for extraneous variables are presented following. Most of these techniques describe what researchers call the **research design** of a study. *Research design* refers to the number of groups in a study, how they are treated, how the individuals are assigned to groups, the number of independent variables, and when the dependent variable is measured.

Control Groups One of the most important practices of experimental designs is the inclusion of a control group. In some designs, the control group is a separate group that receives no treatment or a different treatment than the experimental group but is equal to the experimental group in every other way. The study is set up so that extraneous variables will affect both groups in the same way. A variation on the use of control groups is having participants serve as their own controls by measuring them at least once before and once after the treatment. This is common in single-subject research, which is discussed in more detail later in this chapter. Subjects serve as their own controls in repeated measures group designs, which present two or more treatments to a single group.

Random Assignment of Individuals to Treatments Random assignment of individuals to treatments is a technique used in experimental research to control for possible differences between participants. By randomly assigning persons to treatments, the researcher assumes that many differences in the backgrounds and experiences of the participants will be equally distributed among the groups. However, often in educational research, it is not possible to randomly assign individuals to groups because persons are already in intact groups, such as classrooms. Therefore, other techniques such as those described next are used.

Matching **Matching** is a control technique in which the researcher takes steps to ensure that extraneous variables are equally represented in the experimental and control groups. If Johanna believed that gender might influence her dependent variable—math achievement—she could use a pairwise matching of participants, in which she would match every male in the experimental group with a male in the control group (and do likewise for females). Although this would work effectively for an either/or variable like gender, it would become more complicated if Johanna were concerned about more than two possible values for a variable: for example, a variable like intelligence or achievement that can have a wide range of values. If Johanna were trying to control the extraneous variable of intelligence, she would have to obtain IQ scores on each participant and then do a pairwise matching procedure for IQ. When she assigns a person with an IQ of 120 to the experimental group, she would have to assign a person with an IQ of 120 to the control group. Matching to control multiple extraneous variables or to control for extraneous variables with a wide range of possible values can be a complicated process. This process might eliminate potential participants because there is no individual match for them. This inability to match is especially likely for participants who are high or low in the characteristic that is being matched.

Comparing Homogeneous Subgroups To make the selection process a bit easier while still controlling for extraneous variables, researchers often set up studies in which they are comparing homogeneous subgroups. For example, in a study in which IQ was the extraneous variable, a researcher could propose setting up several IQ groups, each with a range of IQ scores. The use of the Wechsler Intelligence Scale for Children-Revised (WISC-R, a well-known intelligence test) could be proposed to establish a high-IQ group (a score of 115 and above), an average-IQ group (a score of 85 to 114), and a low-IQ group (a score of 60 to 84). Note that these cutoff scores are arbitrary and are based on an average IQ of 100. Although this procedure would facilitate the selection of participants for the study, the generalizability of the findings would be limited to the IQ ranges used for the study, and such limited generalizability would have to be noted as a potential concern by the researcher.

Pretesting of Participants A common type of experimental study, called a **quasi-experimental** study, involves random assignment of whole groups to treatments. To ensure that the groups are similar, researchers often administer a pretest to both groups. A pretest measures whether the experimental and control groups are starting out equal. It is a check of whether there are preexisting differences between the groups in abilities or other characteristics. If there are preexisting differences, then one would not be able to conclude that differences noted at the end of the study are due to the treatment applied. For example, you decide to conduct a study in Elementary School West, which is investigating a new reading program. You obtain permission from the building administrator to conduct your study, but the administrator tells you that only Ms. Johnson's and Mr. Garcia's classes can be used. To conduct a quasi-experimental study in such a situation, the researcher would randomly assign the classes to treatment and control. How? You might flip a coin. If it lands with the head up, Ms. Johnson's class members receive the treatment, and if the tail is up, Mr. Garcia's class is assigned to the treatment group. A pretest would help the researcher to determine if the two groups were equal at the start of the study.

Holding Extraneous Variables Constant Some extraneous variables can be controlled by holding them constant between experimental and control groups. In experimental research in education, the types of variables that are held constant often pertain to the setting or the teachers. If you were comparing two different methods of instruction, you might control for the time of day, the classroom temperature, or the concepts covered by keeping these the same for the treatment and control groups. If two different teachers taught the two methods, you could attempt to control for the quality of the teaching as an extraneous variable by requiring both teachers to have the same number of years of teaching experience (of course, the limitation to this control is the fact that quality of teaching is not solely a matter of years of experience). The lack of control over quality of teaching would be a criticism of the study and should be discussed as a limitation of the study.

Factorial Designs A **factorial design** is one that includes multiple independent variables. Based on the review of the literature, a researcher might identify a potential extraneous variable and decide to build it into the study as an additional independent variable. This means that the final data will be analyzed for possible differences due to each independent variable. These might include variables that can be manipulated by the researcher or that are characteristics of the participants. For example, instead of matching for gender, Johanna might decide to build it into her study as an independent variable. This would allow her to see whether different types of tutoring work better for boys or girls. To further explore the potential effects of both gender and type of tutoring, the researcher could combine these variables into one design with two independent variables: type of tutoring and gender. Table 9.3 shows the groups that would be included in this design.

Statistical Control of Extraneous Variables A common way to deal with extraneous variables is to measure and statistically control them. There are a variety of statistical tests used for this purpose. One statistical procedure used as a mechanism of control is **analysis of covariance** (**ANCOVA**). ANCOVA is defined as a procedure through which participants' posttest scores (scores on the dependent variables obtained after the experimental treatment is given) are statistically adjusted for differences in pretest scores (obtained on the dependent variable before the experimental treatment is given). Sounds like real "massaging of the data," doesn't it? Well, in some sense, that is what ANCOVA is. Assume that you have decided to conduct a research study that examines the effect of a literature-based reading program compared with a program that uses the skills-based method of instruction. You go to a school district and get permission to use Holly Elementary School. Principal Wonderful tells you that you can use Ms. Excited's and Mr. Boring's first-grade classes. No random selection or random assignment of individual

TABLE 9.3 Factorial Design

Computerized Tutoring		Noncomputerized Tutoring
Male	Group 1: Computerized tutoring for males	Group 2: Noncomputerized tutoring for males
Female	Group 3: Computerized tutoring for females	Group 4: Noncomputerized tutoring for females

participants is allowed, so you randomly assign one class to the literature-based instruction and the other to the skills-based instruction. However, being the smart researcher that you are, you are concerned that there might be some initial differences in the students' reading abilities. If the participants are different in reading ability at the beginning of the study, then any difference you might find at the end of the study could not be attributed to your experimental treatment. You decide to test each group before the start of the study, and indeed, you find that Ms. Excited's class has some students reading at the second-grade level, whereas Mr. Boring has a class that is reading pretty much at the first-grade level. Rather than abandon your study because of the differences in the groups, you would use an ANCOVA! The ANCOVA procedure would statistically remove any advantage held by Ms. Excited's class so that the two classes could be compared fairly.

Step 7: Define and Administer the Experimental Treatments

A major part of creating a research proposal is defining the treatments to be administered to the participants, also known as the research plan. This plan should describe all procedures to be used in the study. What happens to the participants in each group? How does the treatment for the experimental group differ from that of the control group? The procedures for an experimental study are described in detail before the beginning of the experiment, and typically, there is little or no deviation from the plan once the study begins. The key to the plan is the differentiation between the experimental and control groups. Keep in mind that detailed descriptions of how both groups are treated must be provided. For example, it is not sufficient to simply say that the control group gets the "traditional approach." That is why Johanna provided details about what would happen to her control group in the study.

Even if the researcher clearly defines the treatment, in practice it might not be delivered in the way the researcher has defined it in the research plan. For example, let's say a school psychologist develops a behavioral plan to be used by a particular classroom teacher. The school psychologist collects data before and after the behavior plan (treatment) is delivered. One concern the school psychologist has is that the teacher may not implement the behavior plan as intended, therefore never really testing the treatment. This would violate what is called **treatment integrity**. *Treatment integrity* is defined as "the extent to which individuals responsible for implementing an intervention can do so as intended by its designers" (National Association of School Psychologists, n.d., ¶3). In order to address this concern, the school psychologist may provide the teacher with a script that clearly defines how the behavior plan should be implemented. This would include things such as appropriate words or phrases she uses in different situations and behaviors to which the plan should be applied.

Defining the treatment is essential because the goal of an experimental study is to "prove" that the experimental treatment has made a difference. So one needs to know what the experimental treatment is and what it is being compared to.

Step 8: Collect and Analyze Data

Using the design of the study, the researcher plans and collects the data. The data produced by experimental studies are quantitative, and as such, the researcher who conducts a study examines the data using statistical procedures. The type of statistical analysis depends on the type of data you collected. As you recall from chapter 4, there are several levels of measurement involved in the collection of data: nominal, ordinal, interval, and ratio. In most quantitative research studies, the researcher first calculates descriptive statistics on the data. Once this has been done, the researcher then selects an inferential statistical test (discussed in chapter 11) that is appropriate for the level of measurement used in the study. Now let's simply examine hypothetical sample data from Johanna's study described in the research vignette at the beginning of this chapter. Johanna compares math achievement of students who are randomly assigned to two treatment conditions: computerized tutoring and noncomputerized tutoring. She concludes at the end of her study that the computerized tutoring group outperformed the noncomputerized tutoring group on a math achievement test. How did she arrive at that conclusion? Let's examine the data from her study in Table 9.4. Note that the scores are from a math test worth 25 points given at the end of the tutoring program.

The data presented in Table 9.4 represents the raw data from the study. Johanna first calculates descriptive statistics on her data. She would calculate a mean to determine the average performance of each group. The mean for the computerized tutoring group is 20.4, while the mean for the noncomputerized tutoring group is 16.5.

What does this suggest? For now, Johanna can only conclude that there was a difference between her two groups. While this provides Johanna with important information, she still needs to use the data to make a decision about her research hypothesis. Is this difference big enough to reject the null hypothesis and support the research hypothesis? Further statistical analysis (inferential statistics) must be performed before that decision is made. This topic is discussed in chapter 11.

Step 9: Make a Decision About the Hypothesis

The data collected and analyzed will provide either support for your research hypothesis or evidence to the contrary. Remember that, in general, the researcher wants to reject the null hypothesis, which states that there is no real difference between the treatment groups—in other words, that any difference found was

TABLE 9.4 Math Scores from Johanna's Research Study

Student Number	Computerized Tutoring Group	Noncomputerized Tutoring Group
1	25	20
2	20	21
3	22	15
4	20	18
5	23	16
6	21	19
7	18	14
8	20	25
9	25	17
10	15	18
11	12	15
12	20	14
13	19	14
14	25	20
15	12	15
16	20	15
17	25	20
18	17	20
19	22	19
20	23	17
21	21	16
22	19	14
23	17	10
24	15	20
25	14	21
26	19	15
27	22	16
28	24	13
29	25	14
30	21	20

due to chance and not to the experimental treatment. Decisions about the null hypothesis are based on probability, which is covered in depth in chapter 11.

Step 10: Formulate Conclusions

The decision made about the hypothesis in an experimental study is based on statistical analysis of a study's data and forms the basis for the conclusions of a study. Confirming a research hypothesis can add to the body of knowledge on a topic and have practical implications. If a researcher finds in a controlled experimental study that one-to-one mentoring decreases student dropout rates, this is an important finding and might support a conclusion or recommendation that schools invest in mentoring programs. Think about how many potential dropouts would stay in school if they received one-to-one mentoring. However, when a research hypothesis is found to be false (that is, the experimental treatment did not make a difference), the conclusion can be equally important. Using the same example, if the null hypothesis about a mentoring program is not rejected (meaning there is no difference between the treatment groups), consider the money a school district might save if mentoring programs did not work—money that could be used to investigate other programs to keep students in school.

Whatever the result of a study, the conclusions must be consistent with the data. Regardless of whether a study produces a supported research hypothesis, the conclusions can still be important to the field of education.

THREATS TO EXPERIMENTAL VALIDITY

Validity in experimental research is generally divided into two concepts: **internal validity** and **external validity**. Internal validity is the degree or extent to which the differences in the dependent variable are due to the experimental manipulation and not some extraneous variable. In other words, at the conclusion of the study, if the two groups are different, are they different because of the treatment? If Ms. Excited's and Mr. Boring's classes are different, what caused the differences? As a researcher, you want the differences to be due to the manipulated variable, in this case, reading instruction. External validity is the degree to which the results are generalizable beyond the sample used for the study. If a difference is found between Ms. Excited's and Mr. Boring's classes (the sample), would the findings be generalizable to other first graders in the district or to other first graders generally? See Box 9.1 for an example of how these concepts were dealt with in an actual study.

BOX 9.1
Internal and External Validity in Action! What Works in the Lab Might Not Work in the Classroom

In an attempt to fulfill the requirements proposed by NCLB (No Child Left Behind Act), a researcher worked to establish a proven practice by conducting a true experimental study. The purpose of this study was to show the effect of a computer-based curriculum on elementary school students' statewide scores in an English Language Arts (ELA) test. To do so, the researcher randomly selected fourth graders from a population of five large urban school districts. Parents of the students all agreed to have their child participate in the study.

Once students were selected for the study, they were randomly assigned to a treatment group and a control group. Students in the treatment group were exposed to the computer-based lessons. These lessons had been carefully developed to address areas in reading for which students in the districts had shown weaknesses on their state ELA assessments. Students participated in these lessons for 30 minutes per day in a computer lab setting where activities were carefully monitored and timed. Students in the control group received instruction on ELA through classroom coursework that their teachers provided. To control for the extraneous variable of time engaging in instruction, the researcher interviewed teachers in the control groups and determined that they, too, spent about 30 minutes per day on ELA.

Students were pretested using a reading literacy instrument. This instrument was analyzed by experts and determined to be aligned with the state assessment. The study took place during the 12 weeks before the ELA assessment. Students in both groups were given a posttest—the same form that they had taken as the pretest measure. Students in the treatment group, who received the computer lessons, had significantly higher gains on the posttest measure than did students who had received the classroom lessons. In addition, students who received the computer-based lessons also outperformed those in the control group on that year's ELA assessment.

The researcher continued to replicate the study, working with other school districts across the state whose students had similar gaps in reaching the state benchmarks

Most of our understanding of these principles is based on the work of Bracht and Glass (1968), Campbell and Stanley (1971), and Cook, Campbell, and Peracchio (1979), who identify the specific threats to external and internal validity. There are seven threats each to internal and external validity. The threats to internal validity are presented in Table 9.5, threats to external validity in Table 9.6.

on the ELA. During replication, the researcher had consistent results with the earlier study. Under NCLB, the computer-based lessons were designated "a proven practice" because experimental research had been used to show the effects of this method. As a result of this, the computer-based lessons were "packaged" and given to middle elementary teachers throughout the state, "guaranteed" to increase student outcomes on the ELA assessment.

Although teachers were eager to adopt and use the computer-based lessons, they found that, for the most part, student scores did not increase. In fact, it appeared that the computer modules had little impact. But wait a minute! How could that be? After all, was not this a proven practice? The answer is yes, under the laboratory-like conditions in which it was studied. However, what the researcher failed to consider was the external validity of the study's findings. The moment that the computer-based modules were no longer administered in a lab in which everything was conducted in a controlled and organized way and implemented into the classroom—a setting that we know does not always go as planned—the study's results were no longer valid or true. For example, we know from the literature and research on teaching and pedagogy that teachers generally do not adopt new curriculum and lessons without modifying them to meet the needs of their students. No matter how good a prescribed curriculum might be, teachers tweak it here and there to make it work for their classrooms. So, one thing that happened to this proven practice is that teachers took it and implemented it in different ways. Many could not spend the entire 30 minutes of class time using the computer-based module, and so they began to "fit it in" a few minutes here and a few minutes there. This did not give students enough exposure to the computer modules to make any difference. In addition, and perhaps the most important aspect, was that most of the school districts did not have a lot of technology. Some of the classrooms had only one computer, and many that did have computers did not have compact disc players or the necessary memory to run the program. Therefore, a method that was shown to be effective in an experimental study, where the conditions were controlled to address threats to the study's internal validity, limited the external validity of the study to the point that it was no longer applicable to the very settings for which it was intended.

Sounds overwhelming? Rest assured that you are not alone in thinking so. Most students as well as professors of educational research moan and groan about learning and teaching the threats to experimental validity. This dismay is due in part to the fact that the concepts are often presented in an abstract form that lacks sufficient relevance to students. Faculty look for ways to make this section interesting and relevant. Even with our collective 35 years of teaching this course, we still struggle with ways to make these threats meaningful to our students. We have embedded these threats in research examples here in an attempt to make the concepts more relevant. Note that each of these threats can be controlled through one of the control techniques listed previously.

Threats to Internal Validity

Researchers should be aware of the threats to internal validity summarized in Table 9.5. Each threat is addressed individually next.

TABLE 9.5 Threats to Internal Validity

Threat	Definition	How to Control
History	Event that occurs outside of the study and affects the dependent variable	Include an appropriate control group
Maturation	A personal change as the result of growth or maturation that can occur in physical, mental, or emotional functioning	Include an appropriate control group
Testing	Pretest of participants on what is being measured on the posttest results in improved scores	Include an appropriate control group
Instrumentation	Instruments used in the study lack reliability, validity, or both	Use reliable or valid instrument(s);train observers and use short observation periods
Statistical regression	The tendency of scores to regress toward the average score, bringing higher scores down and lower scores up	Include a control group selected using same criteria
Differential selection of subjects	Use of already-formed groups that might be different	Use random assignment to groups; pretest participants to see if groups are similar
Mortality	Subject attrition or dropping out	Pretest to obtain information for examining who dropped out of one group and eliminate a similar participant from the other group to maintain equivalent groups; maintain contact with both treatment and control groups

History As part of your safe-school plans, you are conducting a series of training sessions for teachers that are designed to improve teachers' sense of well-being and safety. The training session lasts for 3 months, and as a part of the study, you administer a pretest and a posttest that measure teachers' overall sense of well-being. So far, so good. Let us say that your study is conducted during the shootings that occurred at Columbine (Colorado) High School, where two students went on a shooting rampage that resulted in the deaths of students and teachers. Has your study been influenced by this event? Certainly! Scores on your posttest will likely be lower than they might have been if this tragic event had not occurred. **History** is an event that occurs outside of the procedures planned for the study and affects the dependent variable. Obviously, the longer your study is, the more that history is a potential threat. History does *not* mean the personal history of your participants.

History is especially a problem when studies do not include a control group. Notice that our example refers to only one group getting a pretest and then a posttest. If the study included a control group that did not receive the workshops but did hear about the Columbine tragedy, we could still see whether our workshop helped the teachers who received it because the threat of history was controlled for by keeping it constant between the experimental and control groups.

Maturation As a preschool teacher, you notice that many of your students have difficulty holding a crayon correctly. You decide to work on ways to promote more efficient ways to hold a crayon—a fine-motor skill—so that students will be better able to use a pencil when they get to kindergarten. You set up a mini-experiment in which you evaluate each student's fine-motor skills and then take 5 minutes each morning to "train proper crayon holding." At the end of 3 months, you evaluate the students' fine-motor skills again and find that your treatment worked. Before you give yourself a pat on the back, think about this. Was it your fine-motor "treatment" that made the difference, or was it simply a function of natural change in your students? Perhaps maturation or time improved their performance! **Maturation**, then, is a personal change as the result of growth, which can occur in physical, mental, or emotional functioning. How could you control for maturation? Again, using a control group is one solution; the control group would also be maturing. Because both the treatment and control groups are maturing simultaneously, any differences between groups after the treatment can be attributed to the independent variable.

Testing Many experimental studies involve giving participants a pretest. Let us say that a researcher conducts a study of the effectiveness of a lesson designed to improve students' knowledge about the Vietnam War. The researcher gives students a pretest that assesses their knowledge at the start of the study and proceeds to give them a 2-hour lesson on the war (the treatment). Two days later, students are given the same test, and performance improves. Was it the lesson that improved their performance, or was it simply the short time between test-taking sessions? Did they simply recall the information from the first test on the second test? **Testing threat** refers to *changes in participants' posttest scores that occur because of what they remember from a pretest.* Researchers can control for a testing threat by using a control group; however, a simpler approach might be to simply increase the amount of time between the pretest and posttest to make it less likely that participants will remember the pretest questions.

Instrumentation Not all tests are created equal. Some tests are harder or easier than others. Some are more reliable and valid than others. If your assessments do not accurately measure the variables, the result is an inaccurate assessment of performance. So, even if at the end of your study you demonstrate a difference between groups, it might not be due to the manipulation of the independent variable but simply a function of poor assessments. **Instrumentation** is a threat to internal validity if instruments that lack good reliability and validity are used. Unlike the examples we have given thus far, the validity threat posed by instrumentation is not controlled for by using a control group. To control for instrumentation, make sure you use instruments with good reliability and validity (instruments that have been tested and documented in the literature as valid and reliable measures of your variables of interest). In studies using observers who are recording behaviors, make sure that the observers are unbiased, well trained, and not fatigued by long observation sessions to ensure reliability and validity of observations.

Statistical Regression Assume that you are conducting a study in which you use only students who score in the top and bottom 10% of a class on an algebra pretest. As a result of using extreme scores, you could expect statistical regression (in more common terms, movement of scores) to the mean on an algebra posttest, even if the groups received different treatments. This means that students who scored at the top 10% would regress downward toward the mean, and those who scored in the lower 10% would regress up toward the mean. (If the word *regress* throws you, think of this as a bounce-back effect, bouncing up or down somewhat after the pretest.) At the end of the study, you would not know whether the difference between the groups was due to your treatment or to statistical regression. While it is not common for researchers to conduct studies using the highest- and lowest-score participants in a group, it is possible that studies might examine either only the highest or only the lowest persons. In both cases, statistical regression would apply. Statistical regression, then, is the tendency of scores to move toward the average score, bringing the higher scores down and the lower scores up. How do you control for statistical regression? Yes, control groups to the rescue! If you select participants for both the treatment and control groups in the same way, regression will happen in both groups, and it will be controlled.

Differential Selection of Participants Experimental researchers who randomly assign participants to groups assume that this process controls for possible differences in preexisting abilities. However, random assignment is not always possible due to ethical or practical considerations. When random assignment is not possible, differential selection of participants may be a problem because the researcher is forced to use preexisting groups. Suppose you are conducting a study in which you must use either a convenience sample or cluster random sample because you have to work with established classrooms. In either case, you will be using groups that have been formed before the start of the study. You have no idea how the groups were formed, and in fact, they might be different from each other. Suppose that one class is an honors class and the other class selected is a more heterogeneous mix of students. This would result in the groups being different before they are exposed to the treatment condition. If the honor students outperform the heterogeneous students, what would be the cause of the difference: the treatment or individual differences between the groups (differential selection)? To eliminate the threat of **differential selection** of participants, or using already formed groups that might be different, most researchers give participants a pretest whenever the researchers must use intact groups. Issues with differential selection occur frequently in educational research.

Mortality or Attrition No, this does not mean that your treatment is so severe that some of your subjects die. **Mortality** means participants drop out of your study. The problem with mortality is that you have no control over who drops out. Imagine two groups of students who are mostly equivalent at the start of a study. Then, for some reason, some of the brightest students drop out of the experimental group but not the control group. The groups are no longer equivalent. Once again, if there is a difference at the end of the study, is it due to the treatment or to the fact that the groups are no longer equal due to mortality? Researchers who are planning or conducting long-term studies are particularly concerned about this threat. One way to deal effectively with the risk of participants dropping out and changing the nature of the groups is to use a pretest. If a participant with a particular score drops out of the experimental group, a control group member with a comparable score on the pretest can be eliminated, thereby controlling for mortality or subject attrition.

Mortality may also be a problem in studies using control groups that are compared with a treatment group receiving an intervention program. If persons in the control group are not receiving any communication or attention from the researcher, while the treatment group is in frequent contact due to the intervention, it is likely that mortality will be greater in the control group. It is also likely that highly motivated control group members are the ones most likely to seek help elsewhere and so drop out, making this a significant threat to internal validity. Communication with control group members during long-term studies may be a way to maintain their participation, thereby managing this threat to validity.

Threats to External Validity

Recall that external validity is the degree to which the results of a study are generalizable to populations outside the sample used in the study. The seven threats to external validity are listed in Table 9.6 and are discussed individually next.

Pretest-Treatment Interaction First, pretest-treatment interaction is a problem only when a pretest is used in a study. In some circumstances, the treatment interacts with or sensitizes the participants to the treatment, and the outcome would be different if the participants had not been pretested. When might this occur? Consider the following. You are conducting a study in which the experimental treatment is a workshop designed to improve students' sensitivity toward diversity. Your pretest is a measure of student awareness of their personal attitudes toward diversity. The pretest itself may make students more aware of the issues of diversity and its importance in their lives. If at the end of the study the experimental group is more sensitive to issues of diversity, could your results be generalizable to all other groups receiving the training, or would the findings be generalizable only to groups that received the pretest on attitudes toward diversity? One way to control for pretest-treatment interaction is to use a pretest that does not increase participants' awareness of what behaviors you are trying to change. Some tests disguise their true purpose by using very general titles and subtle questions or items. Professional researchers often use complex designs that include treatment groups that are not pretested to control for pretest-treatment interaction.

Multiple Treatment Interaction Some studies expose participants to multiple treatments that are part of some overarching program or simply expose them to

TABLE 9.6 Threats to External Validity

Threat	Definition	How to Control
Pretest-treatment	Pretest sensitizes participants to the treatment, and outcome would be different if the participants had not been pretested	Include appropriate control groups who also receive pretests and a treatment group that receives no pretest
Multiple treatment interaction	Participants are exposed to multiple treatments that are part of some overarching treatment or simply exposed to more than one treatment; when this occurs, it might be difficult to determine which treatment resulted in any difference that is found, limiting generalizability	Limit the number of treatments delivered, or deliver different treatments at different times; if one treatment has many components, establish comparison groups receiving different components
Selection-treatment interaction	Differences between groups due to lack of random assignment or use of already-formed groups interact with the treatment variable, limiting generalizability to the general population	Pretest groups to see if they differ in any way that might affect treatment effectiveness, or randomly select a representative sample of the population for whom the treatment is intended
Specificity of variables	The more specific the conditions (time, place, participants, tests) are, the more limited the generalizability of the study	Replicate studies with different samples, measures, and settings
Treatment diffusion	Experimental group communicates with control group, providing the latter with information about the experimental treatment; this kind of level of communication might not occur in other studies, thereby limiting the generalizability of the results	Limit communication by conducting the study at different sites or by asking participants not to discuss what occurs until the study is over
Experimenter effects	Researcher may exert unintentional influence on the outcome of the study; if the study was conducted by another researcher, results might be different	Use of researchers who are “blind” or “double blind”; increase researchers’ self-awareness
Reactive arrangement	Simply being part of a study can affect one’s feelings, behaviors and attitudes	Provide information to keep participants from trying to guess the study’s purpose; extend study until reactive effects wear off

more than one treatment. When this occurs, it might be difficult to determine which treatment resulted in any difference that might be found. For example, a researcher conducts a study to determine the effect that attending a charter school has on achievement. In reality, a charter school might include many different treatment components, all of which could affect achievement. For example, a charter school might have more small class sizes and school uniforms. At the end of the study, the charter school students seem to be outperforming their counterparts in regular public school classes. Could the results be generalizable to all charter schools or just charter schools with the same treatment components as the one under investigation? Multiple treatment interactions can be controlled by limiting the number of treatments delivered or delivering different treatments at different times. If one treatment has many components, seek out comparison groups receiving different components. For example, one might compare charter schools that have smaller class sizes with those that have school uniforms.

Selection-Treatment Interaction As we discussed earlier, in the section on internal validity, the threat of differential selection of subjects occurs when already formed groups are used in a study and the groups are different at the start of the study. Any time already formed groups are used or participants are not randomly assigned individually, the generalizability of the study is greatly compromised. In the following example, already formed groups could interact with the treatment variable. You are thrilled to learn that Principal Johnson has given you permission to conduct a research study designed to determine whether cooperative learning improves the math achievement of first graders. However, she tells you that you can use Mr. McDuff's and Ms. White's first-grade classes and cannot randomly form your groups. You have had no control over how Mr. McDuff's and Ms. White's classes were formed; to attempt to manage internal validity, you randomly assign the two classes to treatment and control. Ms. White's class gets the treatment (cooperative learning), and Mr. McDuff's becomes the control group (whole class instruction).

Unbeknownst to you, Ms. White's class contains students who are socially skilled, and they work extremely well on the cooperative learning task. Mr. McDuff's class has students who are not as socially skilled. At the end of the study, the cooperative learning group outperforms the control group. Does this mean that you can generalize your findings to all first graders, including Mr. McDuff's class? Certainly not! If cooperative learning requires a certain level of social skills to work well, it may be effective only for students who have these skills or when their teachers work with them to develop skills as they use the cooperative learning strategy. In this example, an initial difference between the two groups interacted with the treatment to make it more effective, but it is not clear whether the treatment will be effective with groups that do not have good social skills. To control for this, the researcher usually pretests the groups for any initial differences. If the differences between Mr. McDuff's and Ms. White's classes had been identified, the researcher might consider doing the study with two classes with similar social skills.

Specificity of Variables All experimental research is conducted in a specific location, at a specific time, with a specific population, with variables measured with a certain instrument and under a specific set of circumstances. The more specific the conditions are, the more limited the generalizability of the study. A study that is conducted with fourth graders in an inner-city school district, using a specific instructional approach, during the first hour of school, with a specific teacher, and with reading achievement measured with the ABC achievement test may be applicable only to a similar setting. This is one reason why often the purpose of research studies is to replicate previous studies with different groups in different settings using different measures. Any single study has limited generalizability, so it can be useful to replicate studies. However, the criticism of specificity of variables can be avoided by randomly selecting persons and schools that are diverse, using measures that are widely regarded as reliable and valid, and using treatments that can be easily replicated in other settings without specialized resources or circumstances.

Treatment Diffusion Any time you conduct a study with an experimental and a control group, you run the risk of the two groups communicating with each other, thereby "diffusing" or making the treatment less distinct. One of us (M.G.L.) ran into this problem with the second graders she used in her dissertation research on strategy use, memory, and metacognition (knowledge about one's own cognitive strategies and skills). The treatment group received metacognition training in the form of playing a game, and the control group learned to use a simple repetition strategy. Each participant was trained individually, but when some students in the control group participated in the posttest memory task, they used the metacognition strategy (game) that the experimental group had been taught. When the author asked them how they knew to use this strategy, the second graders indicated that their friends (in the treatment group) told them about the game. An interesting outcome, yet not one anticipated by the researcher! *Be aware of the possibility that your experimental group might communicate with your control group, providing the control group with information about the experimental treatment,* and consider how you might design your study to limit or prevent this communication.

Experimenter Effects Good experimental researchers are careful about their own influence on the research they are conducting. Remember that in quantitative research, the researcher maintains an independent and separate role. However, there are times when the researcher may exert unintentional influence on the outcome of the study. These influences can be the result of the personal attributes of the researcher or may occur because the researcher's expectations affect his or her behavior and the performance of the participants. Personal attributes include gender, race, age, or emotional disposition. Researcher influence due to expectations that affect the behavior of research participants (also known as *experimenter bias*) can occur if the researcher, hoping to obtain a difference between the experimental and control groups, gives the experimental group any unintended advantage (more testing time, slower instructions, more attention, positive feedback for correct responses, and so on). Note that bias can occur unintentionally and even unconsciously, especially if researchers have strong expectations regarding which group will do better. When bias is present, generalizability is limited because other researchers may not obtain similar results. To control for experimenter effects, researchers who are in contact with participants are kept **blind** as to the expected outcome of the study or regarding which group is the treatment or control group. The term **double blind** means that the researcher knows neither who is in which group nor what the expected (hypothesized) outcome is. This means that someone must be hired to carry out the study, which may be impractical! It is far more practical to be on guard against researcher effects.

Reactive Effects Have you ever been part of a research study? If so, you know that simply being part of a study can affect your feelings, behavior, and attitudes. One way that reactive arrangement is manifested is through the **Hawthorne effect**. The Hawthorne effect occurs when participants' behavior is affected by their mere participation in the research study and not directly attributable to the treatment. This finding is based on a well-known study conducted at the Hawthorne (and thus the name) Plant of the Western Electric Company in Chicago. The researchers in this study investigated ways to improve the productivity of workers. The independent variable in the study was the amount of light intensity, and the dependent variable was worker productivity. As the amount of light intensity increased, the worker productivity increased. The researchers decided to see what would happen if the light intensity was decreased. Well, guess what? Productivity went up even under conditions of lower light intensity. The workers, aware that they were participating in a research study, increased their productivity regardless of the treatment because they thought the changes indicated that the company cared about them.

Another common reactive effect threat to external validity is called **novelty effect**. Often a new treatment is more effective than an older approach simply because it is new and different. After a while, the novelty wears off, and the new treatment is no better than the older treatment. For example, say that a high school teacher decided to put on a hat each time she was introducing a new concept in class. She uses this treatment in two of her classes (the treatment group) for 2 weeks but not in two other classes randomly assigned to be her control group. She finds that the treatment group shows better understanding of the concepts than the control group. However, the effect may simply show that students pay attention when something new is happening in class. If she continues the hat routine for 2 months, she may find that it is no longer effective. The novelty effect means that a treatment is effective only when it is new or novel and that the treatment's effectiveness will not generalize beyond this initial period of time. In a research study, reactive effects due to novelty are controlled for by extending the period of the study long enough so that any novelty effect will have worn off.

SINGLE-SUBJECT RESEARCH DESIGNS

Single-subject research designs have a long history in education and continue to grow in popularity. In this era of accountability, they allow practitioners and researchers to examine behavior and interventions in a relatively easy manner. The goal of single-subject research designs is to explain and understand how human behavior functions through systematic study. The historical roots of single-subject research are found in the field of experimental and behavioral psychology. Names like Ivan Pavlov, Edward L. Thorndike, John B. Watson, and B. F. Skinner should come to mind. While a discussion of the principles developed by these theorists is beyond the scope of this book, it is important to note that their work formed the basis for **applied behavioral analysis,** which has been associated with a type of true experimental design with a unique twist: the sample size is limited to one participant or a few participants who are treated as one group. It is not uncommon to see these types of designs labeled $n = 1$ designs. In single-subject designs, the participant serves as both the treatment and the control participant. You might be asking, How can one participant serve in both capacities? Good question! The researcher measures participant behavior repeatedly during at *least* two different points in time, when a treatment is not present (the control condition) and again when a treatment is present (the treatment condition).

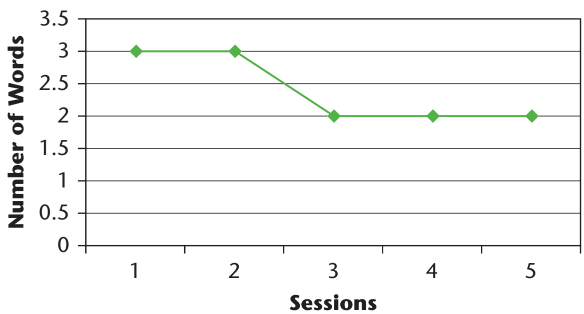
The periods during which the treatment is not present are called **baseline periods**, and the periods during which the treatment is given are called

treatment periods. Baseline periods are essential to single-subject research. The baseline period is used to determine the efficacy of the independent variable (the intervention); therefore careful consideration must go into the collection of baseline data. Baselines allow the researcher to determine the patterns of behavior. Patterns of behavior often include the frequency, duration, and time in between behaviors. Therefore, single-subject researchers use a **repeated measure design.** Repeated measures simply involve the collection of data over time (collection of data points), not only during baseline but during intervention. A good question often asked by students is, How long should baseline data be collected? There is no easy answer. According to Kennedy (2005):

> The goal of a baseline is to establish patterns of behavior to compare to intervention. Therefore, a baseline needs only be long enough to adequately sample this pattern. This judgment depends on the variability of behavior and the patterns of responding. . . . (p. 38)

Kennedy suggests that the baseline should be collected until a clear pattern emerges. Consider the following study. A special educator is conducting a study to determine whether a reward system using praise as a reward (the independent variable or treatment) would increase the number of words (the dependent variable) spoken by a child with autism. Baseline data should be collected until a clear pattern emerges. In the example following (see Figure 9.2), how many data points did the researcher collect to establish a pattern of behavior? If you said five you are correct. After the fifth data point, it became clear to the researcher that a clear pattern emerged from the data. The child used between two and three words across the five baseline sessions.

FIGURE 9.2 Number of Words Spoken by a Child with Autism During Baseline Period



Another feature of collecting baseline data involves examining the data for stability. The more stable the behavior or "steady state" (Kennedy, 2005, p. 38), the more desirable. If the behavior (the dependent variable) is stable prior to intervention, the researcher can more comfortably attribute changes to the intervention (the independent variable).

A critical component of single-subject research is to demonstrate a link or functional relationship between the independent and dependent variable. The researcher hopes to demonstrate that changes in the independent variable will result in changes in the dependent variable. In other words, the dependent variable is *dependent* upon the independent variable. In order to establish this functional relationship, the researcher must control for any extraneous or outside variables that could affect performance of the dependent variable. As you may remember from an earlier discussion, extraneous variables are threats to the internal validity of a study. Given that single-subject research studies are conducted in complex educational environments, these variables may pose a problem for the researcher. In order to minimize the potential threats to internal validity, the researcher uses a variety of experimental designs.

Types of Single-Subject Research Designs

While there are many single-subject designs, the most commonly used designs are **A-B-A designs** and multiple-baseline designs. In all single-subject designs, the A phase of the study represents a series of baseline measurements. The B phase of the study involves the measurements that occur during treatment. In addition, all single-subject designs include continuous measurement of behavior throughout all of the phases.

A-B-A and A-B-A-B Research Designs The simplest single-subject research design is the A-B-A design. In this design, the researcher obtains baseline data (the first A phase), delivers the treatment (the B phase), and after withdrawing the treatment measures baseline data again (the second A phase). Using the example of time on task and rewards as an example, the researcher would collect data on the number of minutes the participant spends on the task to determine a baseline. The treatment phase or rewards would then be given as the child is engaged in a task. The time on task would be measured again through the treatment phase and when the treatment was withdrawn. See Figure 9.3 for an example of how data from an A-B-A design would be graphed.

Some researchers, especially in this example, would use an A-B-A-B design. This designs builds upon the A-B-A design in that following the withdrawal of the treatment (the second A phase), the researcher ends the study with the reintroduction

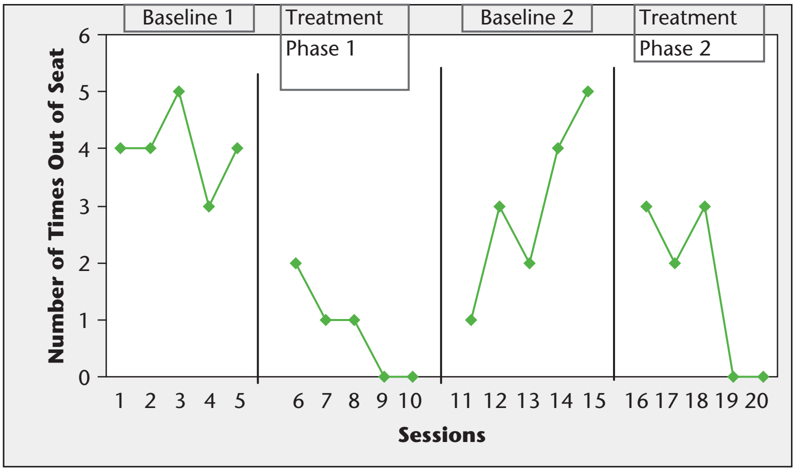
FIGURE 9.3 Sample A-B-A Design

of the intervention or treatment (the second B phase). Under what conditions would a researcher prefer to end with intervention? First, if the treatment was effective, what ethical researcher would consider withdrawing the intervention? Under these conditions it would not be good practice to withdraw it! Furthermore, from a methodological perspective, the A-B-A-B design allows for two separate intervention phases, thereby allowing for replication of the treatment. Replication gives the researcher stronger support for the effectiveness of the intervention. If the behavior (in this case time on task) improves under both treatment phases (e.g., both B phases), the treatment truly works for this particular child.

Let's consider a different example from the one given for the A-B-A design. In this case, the researcher is interested in determining if he can reduce the number of times a child gets out of her seat. In this study, the independent variable or treatment is nonverbal cueing and the dependent variable is the number of times the child gets out of her seat. See Figure 9.4 for an example of a graphed A-B-A-B design.

During the second baseline phase, the behavior returns to about the same level as the first baseline phase. Most researchers would argue that this increases the internal validity of the study. How? When the treatment is removed, the behavior returns to baseline, suggesting that the treatment makes a difference. While this typically happens in an **A-B-A-B design**, it may not always be the case. For some behaviors, the intervention has such an impact on learning that behavior does not return to preintervention levels. When this happens some

FIGURE 9.4 A-B-A-B Design



researchers may add another A-B-A-B phase to the study (yikes . . . this makes the design A-B-A-B-A-B-A-B).

Multiple-Baseline Designs There may be times when researchers simply do not want to withdraw the treatment or intervention phase. This may be the case when there are ethical issues about removing a treatment that is working. In this case, the researcher might choose a multiple-baseline design. **Multiple-baseline designs** are used when the behavior learned during the treatment phase cannot be unlearned (learning to discriminate between the letters O and Q, for example), or when it would be unethical even to attempt to reverse the behavior (a reduction in self-stimulatory behavior of an autistic child). The basic difference between multiple-baseline designs and A-B-A-B designs is that the treatment is never withdrawn. In multiple-baseline designs, the intervention is examined by collecting baseline data on more than one behavior (or person or setting) and applying the treatment sequentially on a time-lagged basis. This means that we begin to apply the treatment to different behaviors (or persons or settings) during different phases of the study. The goal of multiple-baseline designs is to determine whether the treatment is effective across different people, settings, or behaviors. For the sake of developing an in-depth understanding of single-subject research, we have decided to focus our discussion here entirely on multiple baselines across behaviors (remember, they can be conducted across people or settings).

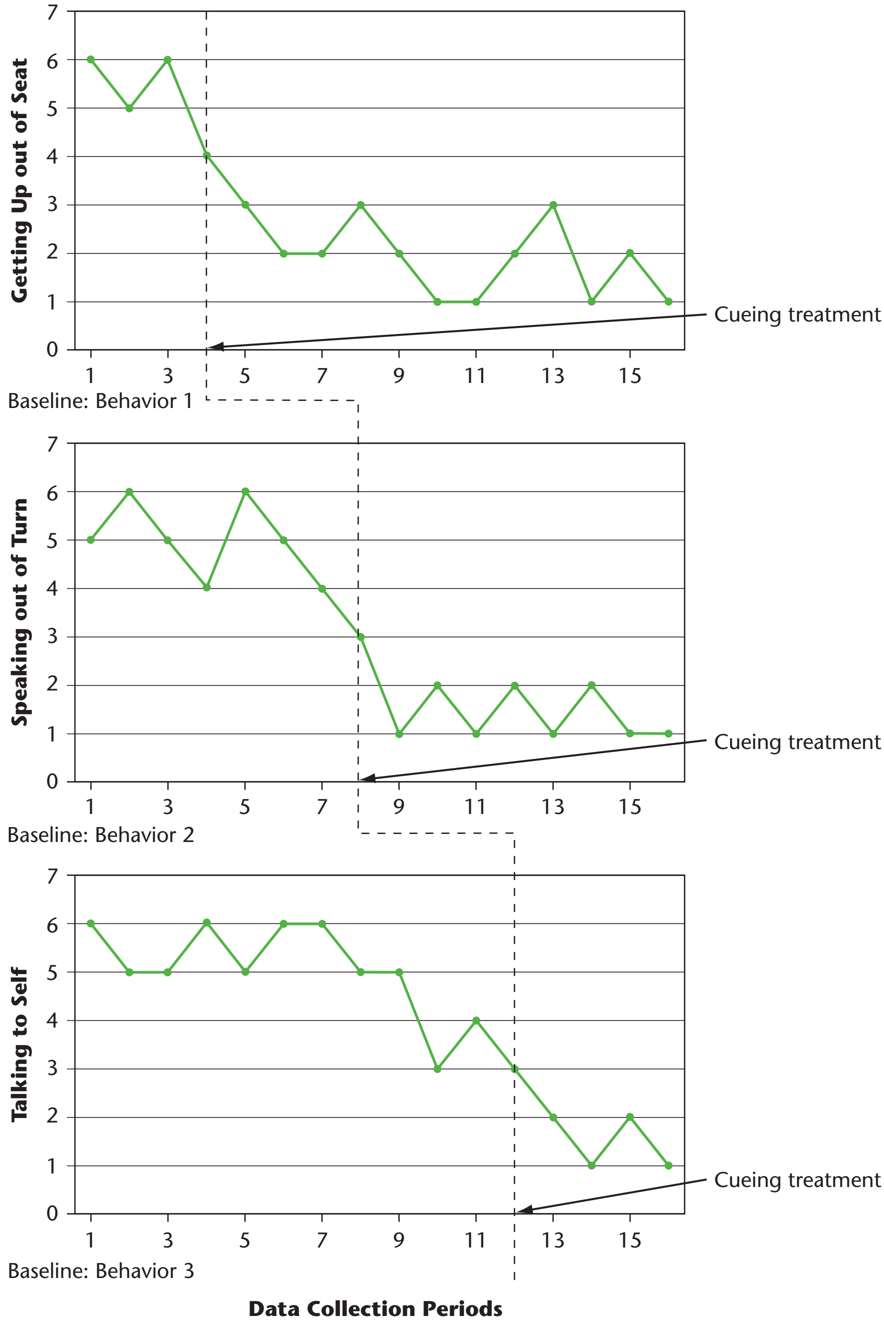
This is how a multiple-baseline study works. First, the researcher identifies at least two behaviors that he or she wants to see changed in a participant. The two behaviors must be different but be similar enough so that a single treatment would be effective. Let us say that a teacher wanted to help a child with severe ADHD learn how to control some disruptive behaviors through the use of a cueing response such as touching her cheek (rather than a spoken reprimand). Our teacher might identify three inappropriate behaviors in a female student: getting up out of her seat, speaking out of turn, and talking to herself. In this design, you would first take baseline measures of all three behaviors four times during the first week. During week 2, the teacher would cue the student for out-of-seat behavior but not for speaking out of turn or talking to oneself. All behaviors would be measured four times again. During week 3, the teacher would cue the student for out-of-seat behavior and for speaking out of turn but not for talking to herself. All behaviors would be measured four times again. During week 4, the teacher would cue the student for out-of-seat behavior, speaking out of turn, and talking to herself. All behaviors would be measured four times again. If the treatment of cueing is effective, each disruptive behavior should change in frequency only when the cueing response is applied to it. Figure 9.5 shows a graph of how the data from this study might look if this treatment were effective.

Whichever type of single-subject design you choose, there are certain rules that any good researcher should follow. Box 9.2 lists these rules.

Internal and External Validity of Single-Subject Designs

As with the group experimental designs discussed earlier, single-subject designs are concerned with issues of internal and external validity. Internal validity (the degree to which changes in the participant occur as a result of the treatment) can be ensured in single-subject research if the researcher carefully obtains baseline data: measurements of behavior over a period of time and before the implementation of the treatment. In our previous example, a researcher decides to determine whether rewards will increase time on task for a child with ADHD. Before the implementation of the treatment—rewards—the researcher measures the participant's time on task behaviors multiple times. It is critical that the baseline data be collected for a sufficient period of time so that the researcher can see either some stability or some trend emerge from the data. In the case of the child with ADHD, the researcher would want to get a sense of how often the child is on task. Observing the child two or three times would not give an accurate measurement of this behavior. Whereas the amount of time on task is likely to vary some from day to day, if a sufficient amount of time is spent observing behavior during the baseline phase of the study, a trend can be determined. How long is a

FIGURE 9.5 Sample Multiple-Baseline Design



BOX 9.2
The Four Rules of Single-Subject Research

Thou shall:

1. Use reliable, nonobtrusive measures. This helps to avoid both a testing effect and pretest sensitization. Observational measures are frequently used.
2. Clearly describe how the treatment is administered and what happens during all time periods.
3. Keep conditions for the baseline and treatment periods as similar as possible to control for extraneous variables.
4. Follow the single-variable rule. Only one variable should be changed from baseline to treatment conditions.

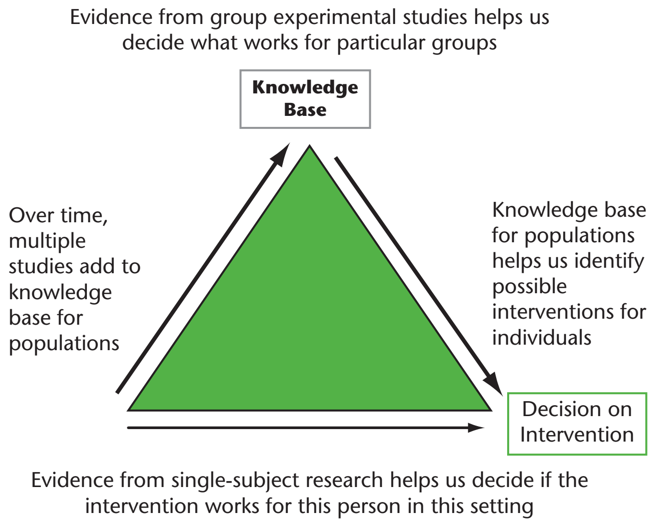
sufficient amount of time? There is no hard and fast rule for this. The researcher must determine, based on his or her observations, when to begin the treatment phase.

A word of caution is warranted here: If the participant's behavior becomes detrimental to his or her well-being, it is better to start the treatment as soon as possible and to forgo the stable baseline phase (even if doing so will threaten the internal validity of the study). In addition, with respect to internal validity, remember that only one variable can be manipulated at a time in a single-subject design. If in the study the researcher was to manipulate not only the reward (one independent variable) but also the lighting in the room (another independent variable) while we were recording the change in time on task behavior, we would not know which independent variable contributed to the change.

Single-subject research is subject to rather low external validity. Why? If the study is conducted on only one person or a small group of persons, realistically the participants will not likely be representative of the population, and therefore, the generalizability of the findings will be low. For single-subject research, then, replicability is the way that a researcher can generalize the findings. If the treatment used in a single-subject study is conducted in other settings with other participants and still produces the same findings, in essence, the findings can be generalized.

This chapter has presented two very different approaches to experimental research: group experimental and single-subject design. Both contribute to our professional knowledge and decision making in different ways, which are summarized in Figure 9.6.

FIGURE 9.6 Group Experimental and Single-Subject Research Add to Professional Knowledge and Decision Making



Source: Adapted from Ylvisaker et al. (2007).

SUMMARY

Experimental research is thought by many to be the only type of research that can suggest true causal relationships. Experimental research is distinctive from other forms of quantitative research in that the researcher controls for or manipulates how groups of participants are treated and then measures how the treatment affects each group. The researcher controls for or manipulates one or more independent variables and examines the effect that the experimental manipulation has on the dependent variable or outcome. The independent variable refers to how participants are treated. In education, the independent variable might be curriculum materials, instructional styles, or specialized trainings, to name a few. The outcome of the study is the dependent variable, which is measured by a test or measuring instrument that produces quantitative data.

Aside from these variables, experimental researchers must consider any potential variables that could influence the groups' performance on the dependent variable. These are called extraneous variables. True experimental research includes studies in which participants are randomly selected and randomly assigned to treatment conditions. Quasi-experimental research involves random assignment of whole groups rather than individuals to treatments. Although not all experimental studies follow the exact same procedure, most follow a standard pattern of practices: selecting a topic, reviewing the literature, defining a research question, developing a research hypothesis, selecting and assigning participants to groups, selecting measurement instruments, analyzing data, and forming conclusions.

There are three types of hypotheses used in experimental research. The directional hypothesis states the direction or the expected outcome of the study, whereas the nondirectional hypothesis simply states that there will be some difference between the variables, but the direction of that difference is not being predicted. The null hypothesis states that there will be no significant differences between the variables after the treatment is applied.

The design of the experimental study is used to control extraneous variables. Design refers to the number of groups in the study, whether the participants are pretested, how they are treated, how the individuals are assigned to groups, the number of independent variables, and when the dependent variables are being measured. After data are collected in an experimental study, they are analyzed to determine whether they support or do not support the hypothesis. Notice that we did not say that the data prove or do not prove the hypothesis. To say so would imply a certain bias on behalf of the researcher.

Another important aspect of an experimental study involves consideration of threats to internal and external validity. Internal validity is the degree to which the difference in the dependent variable is due to the experimental manipulation and that something other than the independent variable is not causing the dependent variable to change. External validity is the degree to which the results are generalizable beyond the sample of the study.

Single-subject designs are a type of experimental design; however, just like the name says, the study is conducted with only one participant or a small number of participants. Single-subject studies typically gather baseline data on the individual before the treatment is administered and continue the collection of data after the treatment is introduced.

KEY CONCEPTS

A-B-A designs
A-B-A-B designs
analysis of covariance (ANCOVA)
baseline periods
blind and double blind
dependent variable
directional hypothesis
external validity
extraneous variables
factorial design
Hawthorne effect
history
independent variable
internal validity
matching
maturation
Hawthorne effect
multiple-baseline designs
nondirectional hypothesis
novelty effect
null hypothesis
operational definition
quasi-experimental study
repeated measure design
research design
research hypothesis
significant difference
treatment integrity
treatment periods

DISCUSSION QUESTIONS OR ACTIVITIES

1. An experimental study has been designed to determine the effectiveness of a new reading program on the reading comprehension of first-grade students. As the researcher, discuss the following:
   - Likely sampling technique
   - Possible research hypotheses (both nondirectional and directional)
   - Extraneous variables that should be considered and ways to control for those variables
   - Ways you might operationally define your dependent variable
   - What you would need to know about the treatment or independent variable
2. In the study mentioned in question 1, assume that you as the researcher have set up the procedure so that one group receives the new reading method and the other group receives the existing method. You decide to pretest both groups to be certain that both groups score about the same on the dependent variable. Each group is then measured with the same posttest. Discuss one threat to internal validity and one threat to external validity. How might you design a study to deal with both of the threats you have described?
3. Design a single-subject research study on a research topic of your choice using each of the following: A-B-A, A-B-A-B, and multiple-baseline designs. Describe the type of baseline you would collect and how often you would collect it before beginning your treatment. Select and justify which design you believe is the best to research your topic.

SAMPLE GROUP EXPERIMENTAL STUDY

McKinney, C. W., & Jones, H. J. (1993). Effects of a children's book and a traditional textbook on fifth grade students' achievement and attitudes toward social studies. *Journal of Research and Development in Education, 27*(1), 56–62.

SUGGESTED READINGS

Sandoval, W. A., & Bell, P. (2004). Design-based research methods for studying learning in context [Special issue]. *Educational Psychologist, 39*(4).

CHAPTER TEN

NONEXPERIMENTAL APPROACHES

Causal-Comparative and Correlational Research

CHAPTER OBJECTIVES

- Define causal-comparative research and differentiate it from experimental research in procedures and ability to provide evidence of causation
- Explain the steps involved in conducting causal-comparative research
- Outline the purpose and characteristics of relationship studies
- Outline the purpose and characteristics of prediction studies
- Summarize the steps for conducting a correlational study
- Define the term statistical significance and the difference between statistical and practical significance
- Describe the types of evidence that support inferences of causality and explain why evidence from correlational studies is weak

RESEARCH VIGNETTE

Rose is a counselor at an urban high school. She frequently talks with students who are working at the same time they attend high school. Although Rose knows that students value their spending money, she is concerned that working might interfere with their studies. She wonders if there is a difference in grades between students who work and those who do not. She decides to collect some data to examine her hypothesis that students who work will have lower grade point averages (GPAs) than students who do not work. When she meets with students, she asks them whether they are employed and how many hours a week they work. After a couple of months, she soon has records on the employment status of more than 100 students. Some students work 5 hours a week or less, but many work 15 hours a week or more. Rose decides to compare students who work 5 hours or fewer with those who work more than 15 hours a week. Knowing that students might differ in more than just their employment status, she decides to select two groups of juniors whose grade point averages for their freshman year were all 3.0 or better. (She assumes that they were not working during their freshman year because few jobs are available to young adolescents.) Rose finds that the students who worked 15 hours or more during their junior or senior years had grade point averages that were on average 0.5 points below those students who worked 5 hours or fewer. She concludes that her hypothesis was supported and begins to discuss her concerns about work and school with her students.

CAUSAL-COMPARATIVE RESEARCH

Like experimental research, causal-comparative research involves comparing groups to see whether some independent variable has caused a change in a dependent variable. Causal-comparative research sets up studies so that possible extraneous variables are controlled. However, the types of research questions addressed in causal-comparative research involve variables that are difficult or impossible to manipulate experimentally, often because they are experiences that have already occurred. Rose's study is one example of a causal-comparative study. Following are some questions that might be addressed using causal-comparative research.

- Do children with a history of abuse have lower levels of academic achievement than children with no history of abuse?
- Do students who are retained a grade have high school graduation rates different from those who are not retained?

- Are women who attend a same-sex college more likely to attain leadership positions after graduation than women who attend coed colleges?

Note that in these questions, we are attempting to see whether one variable (abuse, retention, working, or type of college) causes a change in another variable (academic achievement, graduation rates, or leadership). However, we cannot ethically or practically manipulate the variables that are thought to cause change. Causal-comparative research designs permit the study of the effects of variables that have already occurred or are difficult to manipulate experimentally with human research participants. In many causal-comparative studies, the independent variable—for example, child abuse—has already occurred. This is why the researcher cannot control or manipulate the independent variable; it has already happened.

In other studies, it might be possible to manipulate such variables, but it would be unethical to do so. For example, researchers could not ethically retain one group of research participants for a grade to study the effect of retention on academic performance. Or it may simply be impractical to manipulate the independent variable. If students are already in classrooms with teachers who have established instructional practices, it may not be feasible to randomly assign classes to treatments or individual students to treatments. In this case, a causal-comparative study might be required.

Steps in Causal-Comparative Research

Causal-comparative research often looks deceptively simple. The researcher identifies two groups that had different experiences and then measures how these experiences affected them. However, high-quality causal-comparative research requires careful thinking at each stage. The steps involved in doing causal-comparative research are summarized following.

1. Select a topic.
2. Review the literature.
3. Develop a research hypothesis.
4. Define the independent variable.
5. Select participants and control extraneous variables.
6. Select measurement instruments.
7. Collect data.
8. Analyze data.
9. Interpret the results.

Each of these steps is described following.

Step 1: Select a Topic

In causal-comparative research, the topic is likely to be based on past experiences that are thought to have a strong effect on participants' later behaviors.

Step 2: Review the Literature

The researcher reviews literature to identify what previous research has revealed about the impact of the past experience on later behavior. Potential **extraneous variables** might also be identified through the review of literature. For example, if one was examining the leadership positions of women who attended same-sex versus coed colleges, one might find that students at single-sex colleges tend to come from families with higher levels of income and education. Also, the researcher might find useful information about the methods used to select samples in past studies or measure possible dependent variables. If one wanted to compare children with a history of abuse and those with no history of abuse, previous studies might reveal how researchers were able to identify possible participants. Based on a review of the literature, one would identify an independent variable (prior experience or group difference that cannot or should not be manipulated) and a dependent variable that might be affected by this independent variable.

Step 3: Develop a Research Hypothesis

Research hypotheses for causal-comparative research take a form that is similar to experimental research hypotheses. They both include an independent and dependent variable. The research hypothesis would state the expected causal relationship between the independent and dependent variables. For example, the research hypothesis for the study of working part-time and high school achievement might be

It is hypothesized that students who are employed 15 hours or more a week will have lower achievement than students who are employed 5 hours or less a week.

In this hypothesis, being employed or not employed is the independent variable. The dependent variable is achievement as measured by high school grade point average.

Step 4: Define the Independent Variable

In causal-comparative research, the independent variable describes the different past experiences of the participants. It is important to be clear about the exact differences in the experiences of the two groups being compared. In our opening example, a lot of employment was defined as working 15 hours or more per week. In studying single-sex versus coed schools, one might want to indicate what male-to-female ratios are required for a school to be considered coed. In the study of children with and without history of abuse, one would discuss how the information documenting the abuse was obtained. The definition of the independent variable identifies the two populations from which participants will be selected.

Step 5: Select Participants and Control Extraneous Variables

Unlike participants in experimental research, the participants in causal-comparative research already belong to groups based on their past experiences, and the researcher selects participants from these preexisting groups. An important consideration in designing causal-comparative studies is whether the two groups are similar (comparable) except for the independent variable on which they are being compared. If two groups are formed because they differ on the independent variable, but they also happen to differ on other extraneous variables, the researchers will not know whether group differences on the dependent variable are caused by the independent or extraneous variables. If the employed students were found to have lower scores on a measure of scholastic aptitude, we would have to ask whether their lower academic achievement (the dependent variable) is the result of their employment (independent variable) or their lower academic aptitude (extraneous variable). To rule out the influence of the extraneous variable, the counselor selected groups of students with different levels of employment but with similar aptitudes (based on their freshman grade point averages, presumably before they were employed). Ideally, the two groups should be selected randomly, which Rose did not do. Therefore, she cannot generalize the results of her sample to the whole population of students at her school.

Typically, researchers select participants who differ on the independent variable but are comparable in other ways. Causal-comparative researchers use the same controls for extraneous variables as those controls for extraneous variables (except for random assignment) used in experimental research. These include matching, holding a variable constant, comparing homogeneous subgroups, pretesting (when a researcher is comparing intact groups who are about to receive a treatment that cannot be randomly assigned, such as a new curriculum), use of factorial designs, and statistical controls such as the use of analysis of covariance (ANCOVA) or multiple regression. Although these methods were described in chapter 9, we point out that matching and ANCOVA are especially common in causal-comparative designs because random assignment cannot be used to make sure that participants are similar.

To use these controls, the researcher must obtain measures of the extraneous variables. If a researcher wants to use matching to make sure that the group of participants who have been abused are similar in family income to the group of participants who have not been abused, then information on family income must also be obtained. The most common way that researchers today control extraneous variables in causal-comparative studies is by statistically estimating the effect of the extraneous variable on the dependent variable. Some statistical tests, such as multiple regression, use correlation coefficients to compare the size of effects of the independent variable and extraneous variable on the dependent variable (multiple regression is discussed in more detail later in this chapter).

Another statistical procedure, **analysis of covariance**, or ANCOVA, compares the mean scores of the two groups after the effect of the extraneous variable has been removed. This test estimates how much the extraneous variable affects the dependent variable, and it statistically adjusts the group means to take into account the initial differences between the groups. However, again, to use these statistical controls, there must be a reliable and valid measure of the extraneous variable. Much of the work in designing a high-quality causal-comparative study is focused on measuring and controlling possible extraneous variables.

Note that by itself the random selection of students does not control for extraneous variables that might differ between the two groups. If students working fewer than 5 hours a week generally have parents with higher levels of education than do students working 15 hours or more a week, randomly selecting students may result in samples that differ in parental education level. The random selection ensures that each group is representative of its population. However, if the two populations differ in parental education, so will the two samples randomly selected.

Step 6: Select Measurement Instruments

Selecting appropriate instruments is an important issue in all types of quantitative research. A researcher interested in the question of same-sex versus coed colleges and leadership positions would certainly need to find or develop a measure that accurately measures the dependent variable or types of leadership positions participants had held.

Step 7: Collect Data

In causal-comparative research, there is no treatment to administer. So once the sample and measures have been selected, carrying out the study simply involves obtaining data from the selected participants on the measures. If the measures are archival data, this may involve obtaining permission to access the records, as Rose did in recording GPAs. If a measure involves completion of a questionnaire, procedures must be established to distribute these to the participants and have them returned, or the researcher could administer them in a group setting. Note that obtaining permission or lack of return of the measures might change the sample and open the possibility that extraneous variables have not been controlled.

Step 8: Analyze Data

Data are usually reported as frequencies or means for each group. Inferential statistical tests are used to determine whether the frequencies or means reported for the groups are significantly different from each other. These are the same statistical tests used in experimental research (discussed in chapter 9). Based on the results of these tests, the researcher either accepts or rejects the null hypothesis.

Step 9: Interpret the Results

If the results of the statistical test are significant and extraneous variables have been well controlled, the researcher can conclude that the study provides support for the research hypothesis. However, one should always be cautious about stating that a causal-comparative study has "proved" that a causal relationship exists. Causal-comparative research is valuable in identifying possible causes or effects, but it usually cannot provide definitive support for the hypothesis that one of the variables studied caused the observed differences in the other variable. Evidence from causal-comparative studies is considered to be weaker evidence of causality than experimental studies, which show that a dependent variable changes only after the researcher has manipulated the independent variable. When many causal-comparative studies have been conducted by different researchers working with different samples in different settings and consistent results emerge from these studies, the combined evidence from these studies provides stronger evidence of causality. This has been the case with research on smoking and lung cancer. The probability that these results could occur by chance if smoking does not cause lung cancer is so slight that most scientists who have worked in the area have accepted the combined results as compelling evidence of a causal relationship.

CORRELATIONAL RESEARCH

Is there a relationship between the number of hours that children watch television and their weight? Is there a relationship between the number of teacher absences and student test scores? How can we predict which high school students are most likely to do well in college? What is the relationship among adolescent drug use and alienation from school, having delinquent peers, school failure, and parental use of drugs?

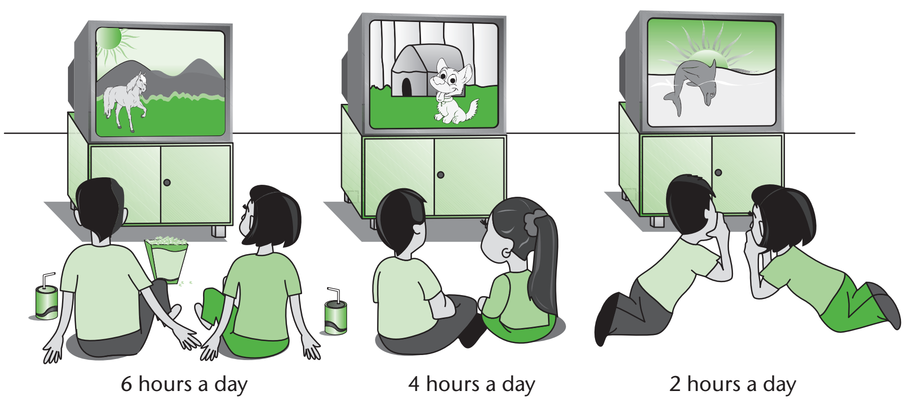
These are examples of research questions that can be explored using correlational research. The purpose of correlational research is to measure two or more variables and examine whether there are relationships among the variables.

First, a cautionary note about what researchers mean when they use the word *relationship* or *correlation*. In research, two variables are said to be "related" when there is an association between the variables such that different amounts or levels of one variable tend to go with different amounts or levels of the other variable in a systematic way. For example, consider how the following illustration (Figure 10.1) suggests that there is a relationship between watching television and body weight.

The drawings suggest that there might be a relationship between the number of hours that people watch television and their weight. The more hours one spends watching television, the higher one's weight is likely to be. In fact, Andersen, Crespo, Bartlett, Cheskin, and Pratt (1998) carried out a study examining just this relationship. They collected data from children between the ages of 8 and 16 years that included the number of hours they watched television and their body mass index (a measure of the amount of body fat).

Note that Andersen et al. (1998) did not claim that watching television directly causes weight gain. One does not absorb calories directly from the television set! However, there is a relationship between watching television and weight, possibly because persons who watch a lot of television eat more food while they view TV or they are less physically active than persons who watch less television. In this

FIGURE 10.1 Television Watching and Body Weight



study, the authors also collected information about the children's participation in physical exercise to see if this variable was related to television watching, weight gain, or both. Many correlational studies do just this, measuring multiple variables and examining the relationships among them. Remember that in saying the variables are related, researchers simply mean that there are patterns in the data that show an association between the variables. Leaping to the conclusion that one variable causes another variable to differ is an inference that is only weakly supported by correlational research. In correlational research, it is best to be tentative when explaining the relationship between two variables, especially if there may be other, related variables that have not been measured (such as exercising or eating habits, as in our earlier example).

Differences Between Causal-Comparative Research and Correlational Research

Like causal-comparative research, correlational research involves quantifying the relationship between two or more variables. However, causal-comparative research is more often used to examine variables that involve dichotomies (such as having a history of abuse or not) or categories (such as single-sex versus coed school). Correlational research is more likely to explore relationships among variables that are continuous.

In addition, causal-comparative research typically involves comparing two groups on one dependent variable. Correlational research often measures many different variables. Although good causal-comparative studies certainly attempt to control many different extraneous variables, the major focus of the study is on whether the two groups differ with respect to the dependent variable. This difference is usually reported in terms of differences between group means rather than with correlation coefficients.

Let us now turn to looking at how correlational research is conducted using two different types of design: a relationship study and a prediction study.

Relationship Studies

One of our graduate students, a teacher in New York City, posed the following research question: "Is the number of days teachers are absent related to student test scores?" He had observed great variability in the number of days teachers were absent both in his own high school and in data for the city as a whole as reported on the State Department of Education Web site. Because scores on the New York State tests varied greatly across the schools, he wondered if there was a relationship between the two variables, teacher absences and student test scores.

His proposal outlined a plan for examining this research question through what is called a *relationship study*.

Relationship studies usually show the following characteristics:

- *Measurement of at least two (but usually more) variables thought to be related.* Typically, relationship studies include many variables because there are few educational or psychological phenomena that are so simple that they can be explained using only two variables. The measured value of variables can vary continuously in small increments (such as the number of days absent) or can be divided into discrete categories such as high, medium, or low (such as reading groups). Variables should be carefully selected based on a review of previous research, with explanations offered for why each variable is included in the study.
- *Data are collected from* one *randomly selected sample of participants.* The sample may consist of individuals, classes, schools, or other groupings (e.g., school districts, states, families), as long as a score for each variable can be determined for each person or group being studied. Persons or groups will also differ in certain characteristics that the researcher is studying. However, the separate units are all considered to be members of one group. So male and female students compose one group of high school students who vary in gender. Tenured teachers are a group that varies in years of teaching experience.
- *Data are collected at one point in time.* Relationship studies appear simple because there are no interventions or programs to set up and no need to observe participants over time. Data are collected "all at once," often by combining the measurement of several variables into long questionnaires that are broken into separate groups of questions to measure the different variables during the data analysis. For example, a questionnaire might measure an adolescent's communication with parents through questions 1 to 10, the extent of their agreement with parental values on questions 11 to 20, and the degree to which their decisions are influenced by discussion with parents on questions 21 to 25. Although the questions are presented all together, the questionnaire includes subscales (groups of questions) that measure different variables. Alternatively, data may be collected using measures from existing records (such as attendance rates, test scores) or from separate questionnaires. However the data are collected, all measures are taken at a single point in time or represent measurements of variables at a single point in time.
- *Scores on each variable are obtained for each individual.* The researcher must obtain a score for each member of the group on each of the variables being examined. Much of the method section in a correlational study discusses how these scores were obtained for all participants in the study.

- *Correlations are computed between the scores for each pair of variables using statistical tests.* The results of correlational tests are used to explain how the variables are related.

Relationship studies often look complex because researchers frequently measure many different variables. Some studies may look at 10 or more variables at once! As the number of variables increases, so does the number of relationships that must be examined. Consider how our student's proposal developed.

He decided to use data from the State Department of Education; therefore, he proposed gathering data about the number of teacher absences for each school and the number of students passing the state test of English language skills at that school from the State Department of Education Web site. (Note that he could have proposed obtaining data on the number of days absent for each individual teacher and the number of students who passed the state tests for each teacher, but he knew these data would be difficult to obtain.) To avoid being overly simplistic, he proposed examining several other variables that were also available on the State Department of Education Web site. He included the number of students receiving free or reduced-cost school lunches as a measure of the number of low-income families. He also included a measure of the number of students for whom English was a second language. So his proposal included four variables: teacher absences, students passing the state test, number of students receiving free lunches, and number of second-language learners. Figure 10.2 shows all the possible combinations of these four variables. Noted in *italics* are the student's predictions that were made as part of his research proposal based on his review of literature.

FIGURE 10.2 Predicted Correlations Among Four Variables

	Number of Students Receiving Free Lunch	Number of Second-Language Learners	Number of Students Passing State Test
Teacher Absences	*Moderately high positive correlation*	*Low or no correlation*	*High negative correlation*
Number Students Receiving Free Lunch	—-	*Low or no correlation*	*Moderate to high negative correlation*
Number of Second-Language Learners	—-	—-	*Low or no correlation*

In a published study (versus a research proposal), the correlation coefficients between each pair of variables would be presented in a table referred to as a **correlational matrix**.

What can one conclude from a relationship study? The safest conclusion is that variables that show large correlations are related. In this case, if our student's predictions were supported, that would mean that as teacher absences increase, the number of students with free lunches also increases (positive correlation). As teacher absences increase, the number of students who pass the state test decreases (negative correlation). As the number of students with free lunches increases, the numbers of students who pass the state test decreases, although to a lesser extent than for teacher absences. Although it may be tempting to conclude that teacher absence or low family income are causes of low test scores, the evidence to support this is weak because there are many other variables that have not been measured or controlled that might also contribute to low test scores. Perhaps teachers who are absent frequently are poor teachers. (Their absence may indicate a lack of commitment to teaching.) It may be the poor teaching, not teacher absence itself, that causes low test scores. Relationship studies are often referred to as "exploratory" because they suggest relationships between variables that need additional study before claims of cause-and-effect relationships can be made.

Prediction Studies

A prediction study is another design used in correlational research. Like relationship studies, **prediction studies** examine correlations between variables, but the goal is to identify one or more variables that can predict changes in another variable measured at a later point in time. For example, Proctor, Moore, and Gao (2003) used a prediction study to examine the relationship between television viewing and weight by looking at correlations between daily hours of television viewing during childhood and body fat during adolescence. In this case, television viewing during childhood predicted the amount of body fat during adolescence. In a prediction study, the variable that is measured at the earlier point in time is called the **predictor variable** because it is used to predict something occurring later. The variable that is being predicted is called the **criterion variable**. In Proctor et al.'s study, the hours of television viewing in childhood was the predictor variable and body fat during adolescence was the criterion variable.

The characteristics of prediction studies are listed below, and in many respects, they are similar to the characteristics of relationship studies. The major difference is that the behavior or experience measured by predictor variables occurs before the behaviors or experiences represented by criterion variables.

- *One or more predictor variables and one or more criterion variables are identified.* Based on a review of literature on the topic, the researcher specifies in advance which variables are considered predictor variables and which are considered to be criterion variables.
- *Data are collected from one group of participants to measure variables representing two different points in time.* The behaviors or experiences represented by the predictor variables must precede the behaviors or experiences represented by the criterion variables. In many studies, the collection of data on the predictor variables precedes collection of data on the criterion variables. However, predictor variables may be measured using records or self-reports of behaviors that happened earlier. In this case, data could be collected for both the predictor and criterion variables at the same time, but the data represent behaviors or experiences that occurred at different points. For example, student grades in high school might be collected from existing records to see whether these predict grade point averages in college. Both the high school and college grades would be collected at the same time.
- *Data are collected from one group of participants.* As in relationship studies, prediction studies collect measures of all variables on a single group of randomly selected participants.
- *Scores on each variable are obtained for each individual.* Because variables in prediction studies are often measured at two points in time, prediction studies may involve data collection procedures and methods for staying in touch with participants over time.
- *Correlations are computed between the predictor variables and the criterion variables.* The results of correlational tests are used to examine whether there is a significant relationship between the predictor and criterion variables. If a strong relationship exists between a predictor variable and a criterion variable, the predictor variable can be used to make predictions regarding the criterion variable. This is done by statistically generating a prediction line similar to that shown in chapter 3 (refer back to Figure 3.10). Correlational tests may also examine whether several predictor variables can be combined to yield a more accurate prediction of the criterion variable. In the example discussed earlier, in addition to high school grades, colleges might use Scholastic Assessment Test (SAT) scores and other achievement test scores to predict student grade point averages.

Steps in Conducting a Correlational Study

The steps for conducting either a relationship or a prediction study are presented next.

1. Review of the literature to identify variables.
2. Select a sample.
3. Identify a measure for each variable.
4. Collect data.
5. Analyze data.
6. Interpret the results.

Step 1: Review the Literature to Identify Variables

Past literature is examined to identify variables other researchers have found to be related. Often new correlational research seeks to expand the understanding of complex phenomena by selecting new variables or new combinations of variables and examining the relationships among them. Assume that a review of literature yielded one study that found that underaged drinking was negatively correlated with parental monitoring and a separate study showing that underaged drinking was positively correlated with peer acceptance of drinking. A proposed new study might examine all three of these variables within one group to see whether peer attitudes or parental monitoring was more strongly related to youth drinking.

Step 2: Select a Sample

Ideally, individuals should be selected randomly from a larger population. The sample should contain at least 30 individuals (or separate entities, such as schools in our example of teacher absenteeism and student test scores). Larger samples will increase the generalizability of the results.

Step 3: Identify a Measure for Each Variable

The most complex part of doing a correlational study is finding or developing valid measures for the variables. For example, if you are interested in self-esteem as a variable, how is that translated into a number? Care must be taken to ensure that the measures are reliable and valid.

Step 4: Collect Data

Data for each variable must be obtained for each person in the study. Because pairs of scores will be correlated, it is important to have a way to link data from different measures so that scores belonging to the same person can be grouped together. This may be particularly difficult in prediction studies in which data are collected at different points in time. To preserve confidentiality, scores for each person on the predictor variable are usually identified with a number code that is linked to a list of names kept in a locked file. This list is used only to code data collected later on the criterion variable so that scores can be grouped. See Box 10.1 for an example of a researcher who maintains confidentiality.

Step 5: Analyze Data

The test used will depend on whether the data consist of variables that are **continuous** (changing in small increments, such as for test scores) or **dichotomous** (separated into distinct categories, such as gender or tenured and nontenured). Table 10.1 lists the most commonly used statistical tests for examining relationships between pairs of variables.

BOX 10.1
Ethical Issues in Correlational Prediction Studies

Correlational prediction studies examine whether a measure taken at one time can predict a later behavior. To do this, researchers must be able to link the two measures taken by each participant. To protect confidentiality, however, no identifying information may be kept with the measures. How can one preserve confidentiality over a long period of time and still manage to connect each person's earlier score to their later score? For example, if a researcher administers a questionnaire on career aspirations to children in eighth grade and later collects information about their final high school grade point averages, how can she or he know which career questionnaire should be paired with which high school grade point average?

One way that researchers do this is by numbering the questionnaire measures often used in correlational research and attaching a removable sheet that includes the person's name and the number on his or her questionnaire. The sheets with identifying information are removed either before the person completes the questionnaire or shortly thereafter. A list is constructed of the persons' names and numbers. This list is kept in a locked file, and no one except the head researcher is allowed to access it. All data are tabulated anonymously, usually by persons who cannot access the list. When the second measure is taken later, the list is used to pair up each person's two measures. After the data have been paired and entered into the computer, the list is destroyed.

TABLE 10.1 Common Statistical Tests for Examining Relationships Between Variables

Statistical Test	Symbol Used for Correlation Coefficient	Type of Variables and Remarks
Pearson-product moment correlation	r	Two continuous variables and samples of 30 or more; most stable test with smallest amount of error
Spearman rho correlation	ρ	Data reported as ranks or continuous variables in studies with samples smaller than 30
Point-biserial correlation	r_{pbis}	Used when one variable is continuous and the other is dichotomous
Phi coefficient	π	Used when both variables are dichotomous
Eta or correlation ratio	η	Used when relationship is expected to be nonlinear

Step 6: Interpret the Results

Both the size and strength of correlation coefficients are considered in interpreting results. Issues relating to statistical versus practical significance of correlations are discussed in the next section of the chapter.

Evaluating Correlational Studies

When evaluating correlational studies, researchers take several issues into consideration: the sampling techniques, the reliability and validity of the measurements, whether the results are of statistical or practical importance, and how carefully the variables are selected.

Sampling Techniques In chapter 8, we discussed the use of random sampling techniques to select a sample in survey research. Like other types of quantitative research, the goal of correlational studies is to generalize results to a larger population. To make valid generalizations, random sampling techniques must be used in selecting a sample. For example, to determine whether SAT scores accurately predict freshman grade point averages, one would need to select a representative sample of students applying to colleges. If studies do not use random samples, researchers must exercise caution in generalizing the results and conclusions of the study.

The size and heterogeneity of a sample may also affect the results of a correlational study. In general, wide variability in scores is helpful in a correlational study because it allows one to see the patterns among the variables for high, middle, and low scores. If scores on a variable for a particular sample are too homogeneous, the data are said to have restriction of range, meaning that the full range of possible scores is not represented. Restriction of range may mask the true relationships between variables in a correlational study. Consider the scatterplot that shows the relationship between weight in pounds and physical exercise (top of Figure 10.3). What would happen if our sample was restricted to only children who exercised 30 or more minutes a day? The bottom of Figure 10.3 shows this scatterplot with the restricted range on level of exercise.

The scatterplot shows that when we restrict the range of children to those who exercise 30 minutes or more a day, the correlation between weight and exercise appears to be much weaker than it did in the top graph of Figure 10.3. In a correlational study, the best way to see the true relationship among the variables is to have samples that are heterogeneous, with a wide range of scores on the variables of interest.

Reliability and Validity of Measurement

Because correlational research involves the measurement of many different variables, the quality of the measures used is a major consideration. As discussed in chapter 4, reliability and validity are two concepts that are used to judge the quality of educational measures. *Reliability* refers to the consistency of measurement. *Validity* refers to whether the measure accurately and appropriately measures whatever it is supposed to measure. Correlational studies should always discuss evidence for the reliability and validity of their measures. If no evidence of reliability or validity is provided or available, the results of the study may be suspect. For example, if the children exaggerated the amount of time that they exercised, the data would not accurately represent the true relationship between weight and exercise.

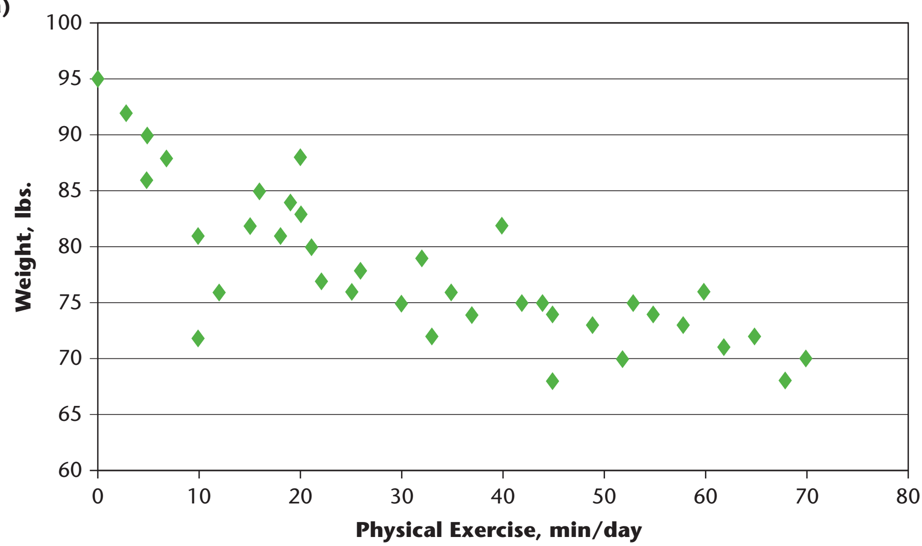
Statistical and Practical Importance

As noted earlier, correlational studies often report relationships among many different variables. For example, Figure 10.4 shows results from a study on bullying that included seven variables examining types of bullying and victimization (being bullied) and perceptions of school safety. In the correlation matrix, each variable is numbered, and each cell in the table shows the correlation between two variables.

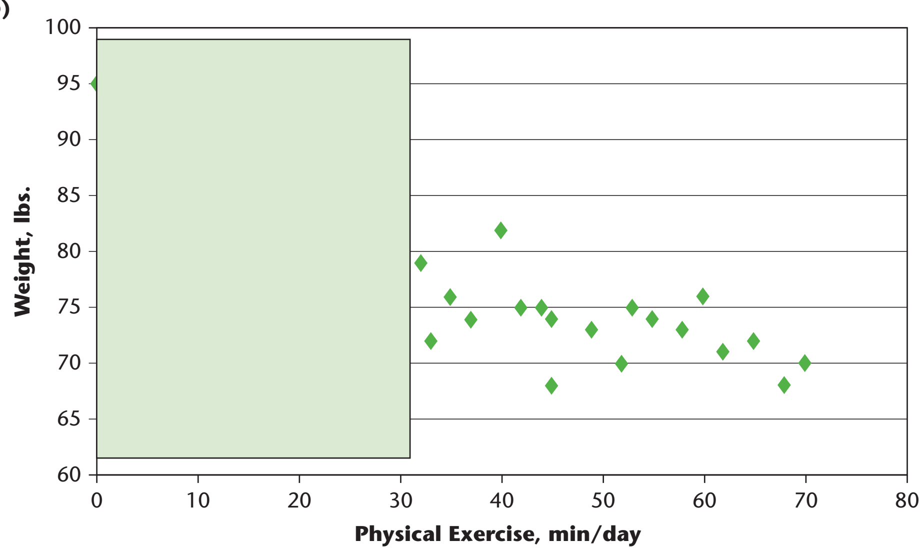
The double asterisks asterisks (**) in Figure 10.4 indicate which of the correlations are statistically significant. A correlation is considered statistically significant

FIGURE 10.3 Scatterplot Showing Restriction of Range in Amount of Physical Exercise

(a)



(b)



if the probability that the correlation was obtained due to chance is less than .05. This probability is usually noted on a correlational matrix using the notation $p<.05$. In Figure 10.4, "**$p<.01$" at the bottom of the matrix indicates that the probability of correlations labeled with a double asterisk being due to chance is less than .01 and therefore, statistically significant. If a correlation is not statistically significant, it means that any relationship observed between the variables could have occurred due to chance. Researchers would assume that a nonsignificant correlation coefficient indicates that the variables are not truly related.

The *p* values listed at the bottom of a correlational matrix or following a correlation coefficient are referred to as the **significance level** of a correlation. A small *p* value (e.g., $p<.01$ is smaller than $p<.05$) means that it is unlikely that the correlation could have been due to chance. Researchers assume that a smaller *p* value means that they can have greater confidence that there is a true relationship between the variables. However, a small *p* value does not mean that the correlation is practically important. Even very small correlations can be statistically significant

FIGURE 10.4 Correlation Matrix for Study of Bullying

	2	3	4	5	6	7
1. Been in a physical fight	+0.35**	+0.33**	+0.23**	+0.33**	+0.32**	−0.16**
2. Used a weapon	—	+0.54**	+0.19**	+0.50**	+0.32**	−0.19**
3. Been arrested	—	—	+0.13**	+0.41**	+0.22**	−0.12**
4. Been teased because of race or gender	—	—	—	+0.33**	+0.38**	0
5. Been threatened with a weapon	—	—	—	—	+0.42**	−0.22**
6. Had property stolen or damaged	—	—	—	—	—	−0.22**
7. Perceived school safety	—	—	—	—	—	—

Note: $N = 3{,}542$; **$p<.01$.

Source: Go & Murdock (2003).

if a large sample is used. Consider that in the bullying example, even a small correlation of –0.12 (between "Been arrested" and "Perceived school safety") was significant at $p<.01$. In fact, all 21 correlations in this study were reported as statistically significant at $p<.01$. One reason that this occurred here is this study used a large sample of more than 3,000 persons. With large samples, even very small correlations are likely to be statistically significant. However, is there any practical importance to knowing that two variables have a weak relationship?

The practical importance of a correlation depends on the size of the correlation coefficient and the decision that will be made based on the correlation. Slight to low-moderate correlations may be useful to researchers who are trying to build models of the relationships among many different variables. However, such correlations would not be accurate in making predictions about the performance of a group. So a college looking for a test that can predict the overall performance of their freshman class would want one that research had revealed to have a moderately strong or strong relationship with freshman grade point average. For individual predictions, such as those made by a college admissions committee, strong or very strong correlations would be required. Table 10.2 summarizes the

TABLE 10.2 Practical Interpretations of Correlation Coefficients

Size of Correlation	Strength of Relationship	Possible Uses or Interpretations
0 to 0.19	No relationship or weak relationship; likely to be statistically significant only in large samples of 1,000 or more	If statistically significant, may show weak relationship between variables, but no practical use in either relationship or prediction studies
0.20 to 0.34	Slight relationship detectable in samples of 100 or more	Useful in examining relationships among variables but not accurate for group or individual predictions
0.35 to 0.64	Moderately strong relationship	Typical range of correlation coefficients for relationship studies; may be useful for group predictions
0.65 to 0.84	Strong relationship	In relationship study, shows a strong association; accurate prediction for groups possible; at the high end of this range, accurate individual prediction possible
0.84 or greater	Very strong relationship	Correlations this high are rarely seen in relationship studies and would indicate that the two measures are measuring the same variable; accurate individual prediction is possible; studies of reliability or validity would also seek relationships this strong

interpretations and practical uses of correlation coefficients of different sizes in educational research, based on work by Cohen and Manion (1994).

Careful Selection of Variables

As noted earlier, both relationship and prediction studies may include a large number of variables. The variables in a study should be carefully selected based on previous research. Researchers should avoid what is sometimes called a "shotgun approach," in which any and all variables that the researcher can think of are included in the study. The shotgun approach is a poor research practice because correlational tests examine relationships between each pair of variables. Therefore, including a large number of variables means that many different correlation coefficients must be computed to examine the relationship between each pair of variables. It is likely that some of these correlation coefficients will be significant due to chance. If a significance level of .05 is used, then 5 times out of 100 one may get a significant correlation due to chance. If all combinations of seven variables are tested, 21 correlation coefficients will be computed (as in the bullying study discussed previously). It is likely that at least one of these correlations was significant due to chance. However, there is no way to tell which of the correlations were due to chance and which represent true relationships between variables.

In addition, including a large number of variables involves extra work for both the researcher and possibly the participants. The researcher must find additional reliable and valid measures for each of the variables, and the participants may then have additional questions to complete on the measures. However, including multiple variables is a good research practice if one has reason based on previous research to expect that the variables will be related.

Correlation and Causation

Traditionally, researchers have been hesitant to infer that a correlation between two variables indicates a causal relationship. Traditional research wisdom admonishes, "Correlation does not mean causation." Many older educational research books insist that it is never appropriate to infer causation from a correlation, and indeed, there is good reason to exercise caution when thinking about whether a correlation indicates that one variable causes another to change. Consider the following correlation:

> *There is a positive correlation between average outside temperature and the number of children with broken bones seen at hospitals.*

Does this finding mean that warm weather causes broken bones? Of course not! During warm weather, children are more likely to play outside in active pursuits and fall down, and this is what causes the broken bones. In correlational research, playing outside would be considered an **intervening variable**: the variable coming between air temperature and broken bones that is the true causal variable (Figure 10.5).

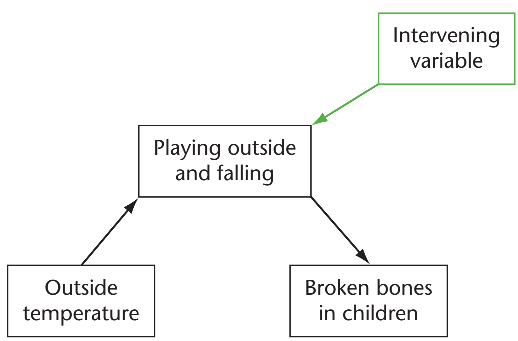
There is a positive correlation between the amount of disciplinary control parents use and a child's behavior problems.

Does the amount of disciplinary control cause the child's behavior problems? Here it is more tempting to infer that there is a causal relationship, because other research shows that parenting that involves discussion and reasoning with children is correlated with fewer behavior problems. However, we know that parents also modify their parenting based on their child's behavior and personality. In other words, the child's behavior problems may cause the parents to use more or less control in their parenting. The direction of causation works in two ways: parenting may partly cause behavior problems, but children's behavior problems also cause parents to change their parenting.

If the examples presented here have not convinced you that correlation is not sufficient to prove causation, consider a study found by one of our graduate students. This study reported that parental involvement in homework was negatively correlated with student achievement! Does this mean that parental involvement in homework causes students to do more poorly in school? What other explanations might you give for the correlation? The study explained that parents of students who had difficulty in school spent more time helping their children. If students were doing well, the parents did not need to provide help. Again, the direction of the causation was not clear based on just the existence of a correlation.

As we noted earlier, many older educational research books insist that it is never appropriate to infer causation from a correlation. However, whether a correlation can be used as evidence of a causal relationship depends on both the design of the study and the type of statistical test used to analyze the results. Box 10.2 discusses the types of evidence required to show that one variable causes another to change within correlational, causal-comparative, and experimental research.

FIGURE 10.5 Example of Intervening Variable



BOX 10.2
Inferences About Causation

What does it mean to say that one thing causes another to change? Common sense suggests that it is easy to tell if one thing causes a change in another. Take, for example, the observation that during the summer, temperatures are hotter. Many people assume that the Earth must be closer to the sun during the summer because, usually, the closer a hot object such as the sun is to you, the warmer it makes you feel. However, in the Northern Hemisphere, the sun is actually farther away from the Earth during the summer than during the winter. It is the angle of the sun's rays (which are more direct during summer) that causes hotter days. So, in this case, common sense is not sufficient. One must be careful about the type of evidence presented to support inferences of causality.

Researchers have identified several levels of evidence that can be used to support inferences about whether one variable truly causes a change in another variable. The types of evidence are listed following in order of increasing strength. Note that each higher level also assumes that the lower levels are true.

- *Correlation. One variable occurs in association with another variable.* If two variables occur together frequently and covary in systematic ways (e.g., are correlated), it is possible that one is causing the other to change. However, this type of evidence is considered weak.
- *Timing. One variable precedes another.* Because change happens over time, a cause must precede the effect that it controls. Therefore, if experiences represented by a variable occur before behaviors measured by another variable, evidence that the experience caused the change in the behavior can be inferred. For example, in the study noted earlier, TV watching in childhood preceded obesity in adolescence. This is stronger evidence than simply finding a correlation between TV watching in childhood and childhood obesity measured at the same point in time. However, it is possible that other variables were also changing at that time. Unless these other variables are controlled (see following), one cannot be certain which variable produced the change in behavior.
- *Control of other possible variables.* If other possible variables can be controlled, evidence of causation is strengthened. However, it is usually not possible to know whether all other possible causal variables have been controlled. Correlational research attempts to control other possible variables through measurement of multiple variables and the use of statistical controls. (See discussion of multiple regression in next section of chapter.) In causal-comparative and experimental research, such variables are called extraneous variables and are controlled through a variety of statistical and procedural methods.
- *Direct manipulation of one variable followed by a change in another variable.* The strongest evidence of causation occurs when one directly manipulates or changes one variable while controlling as many other variables as possible. This is the method used in experimental studies.

Applying the information in Box 10.2, one would conclude that prediction studies offer stronger evidence of causation than relationship studies, although this evidence is not as strong as the evidence provided in experimental studies. The next section of the chapter discusses one sophisticated statistical test—multiple regression—that is used in correlational studies to examine correlations while controlling for other variables. Many researchers now consider this technique a viable way to provide evidence of causation in correlational studies.

MULTIPLE REGRESSION STUDIES

The statistical tests discussed earlier were designed to examine relationships between pairs of variables. However, as understanding of the issues in education has progressed, researchers have looked for ways to examine relationships among multiple variables. Multivariate correlational statistics provide a way to examine multiple variables at once and to separate the contributions of different variables.

Purpose of Multiple Regression Studies

Multiple regression is used for two purposes: first, it allows examination of relationships between two variables after the influence of other variables has been removed. Second, it allows examination of how accurately a combination of several variables can predict a criterion variable. For example, say you wanted to predict first-grade students' performance in learning to read. You hypothesize that several variables might predict later reading achievement: the number of hours per week that parents read to their child, the child's phonemic awareness of the sounds, the number of children's books in the home (a measure of the literacy environment), and the child's interest in reading. Assume that you have reliable and valid measures of each of these predictor variables for a group of first graders and can obtain a measure of their reading scores at the end of first grade (the criterion variable). However, you know that parents who read to their children are also likely to have books at home for the children. These predictors will overlap in their contributions to final reading scores, although not completely because some parents may be able to afford more books than others or may make up for lack of time to read with their child by buying books. So it is helpful to be able to separate the contribution of reading to a child from that of having books at home.

Multiple regression provides a way to do this by calculating correlation coefficients, referred to as **beta weights,** for each predictor variable. The beta weight indicates the relationship between the predictor variable and the criterion after the effects of all other predictor variables have been statistically removed. Beta weights are similar to simple correlation coefficients in that they measure the association between two variables; however, they also isolate the contribution made by each predictor variable by statistically controlling the overlap of that variable with other predictor variables. The beta weight allows you to see the relationship of reading to the child to the criterion variable (reading scores) separate from that of the number of books at home.

A multiple regression also allows you to see how several predictor variables together might improve a prediction. **Coefficients of multiple regression**, symbolized as R, are calculated to show how much the correlation increases as each predictor variable is added to the analysis. Results of a possible multiple regression analysis for the study of reading are presented in Table 10.3.

In the table, the largest beta weight is 0.20, and it indicates that phonemic awareness by itself is the best predictor of reading performance. The next most accurate predictor is parent reads to child. Of these four predictors listed, the child's interest in reading is the least powerful predictor, as shown by the lowest beta weight of 0.11. In the third column, 0.20 again shows the correlation of phonemic awareness and reading performance. The next coefficient, 0.30, shows the correlation between phonemic awareness combined with parent reads to child and reading performance. Similarly, the coefficient of 0.35 shows the correlation between reading performance and the three combined predictors: phonemic awareness, parent reads to child, and number of books at home. Finally, the coefficient of 0.38 indicates the correlation of the combination of all four predictor variables with reading performance. Note that the accuracy of the prediction goes up (shown by the increasing value of R) as predictor variables are added. The coefficient of multiple regression, R, indicates whether the addition of each predictor variable significantly improves the accuracy of the prediction. In this case, all predictors except the child's interest show a statistically significant improvement in reading performance, as shown in Table 10.3 by the * indicating that $p<.05$.

Multiple regression is becoming the most popular statistical test used in correlational research. It is used in causal-comparative and experimental research as well, because it allows a more realistic picture of the many different variables that comprise complex educational phenomena.

TABLE 10.3 Multiple Regression Analysis for Study of Reading

Predictor Variable	Beta Weights	R Coefficient of Multiple Regression
Phonemic awareness	0.20	0.20*
Parent reads to child	0.18	0.30*
Number of books at home	0.16	0.35*
Child's interest in reading	0.11	0.38

*$p<.05$

SUMMARY

Like experimental research, causal-comparative research involves comparing two groups to see whether some independent variable has caused a change in a dependent variable. Unlike experimental research, however, causal-comparative research is able to investigate research questions in which the independent variable cannot be manipulated because of ethical concerns or because the variable has already occurred. Socioeconomic status, preschool attendance, and number of siblings are examples of variables that cannot be manipulated.

There are ass series of steps a researcher must follow to conduct a quality causal-comparative study. As in other types of cause-and-effect studies, the researcher selects a topic, conducts a thorough review of the literature, identifies an independent variable, and carefully defines it. Next, the researcher selects two preestablished groups to study: one group implementing the treatment or independent variable of interest to the researcher, and the second group using the other method. The researcher also attempts to control for extraneous variables. At the end of the two treatments, the researcher collects data from the participants using a reliable and valid measurement tool and compares the results of both groups on the measure to determine whether a significant difference was found. If such a difference between the two groups is found, then the difference is attributed back to the independent variable.

Unlike causal-comparative research, correlational research examines relationships and looks to see whether there is an association between the variables such that different amounts or levels of one variable tend to go with different amounts or levels of the other variable in a systematic way. Causal-comparative research is often used to examine variables that involve dichotomies (such as having a history of abuse or not) or categories (such as single-sex versus coed school). Correlational research is more likely to explore relationships among variables that are continuous. In addition, causal-comparative research typically involves comparing two groups on one dependent variable. Correlational research often measures many different variables.

To conduct a correlational study, the researcher must have two variables (or scores) for each participant. There are two types of correlational studies: relationship and predictive studies. Relationship studies seek to determine whether a relationship exists between two or more variables, whereas the purpose of predictive studies is to identify one or more variables that predict the results of participants on another variable. In published studies, the correlation coefficients between each pair of variables are presented in a table referred to as a correlational matrix.

In analyzing their data, correlational researchers need to identify the type of data and select the appropriate statistical analysis. Like other types of quantitative research, the goal of correlational studies is to generalize their results to a larger population. To make valid generalizations, random sampling techniques must be used in selecting a sample. A minimum of 30 participants is required for a correlational study. The size and heterogeneity of a sample may affect the results of a correlational study. Measurement tools used to collect data for the variables in a correlational study must be reliable and valid.

Ultimately, the researcher wants to conclude that there is a statistically significant relationship between the variables that is not due to chance. A correlation is considered statistically significant if the probability that the correlation was obtained due to chance is less than .05. This probability is usually noted on a correlational matrix using the notation $p<.05$. Multiple regression is another approach to correlational research that moves our understanding from one that is a relationship between variables to that of an inference of causality.

KEY CONCEPTS

analysis of covariance
beta weights
coefficient of multiple regression
continuous and dichotomous variables
correlational matrix
criterion variable
extraneous variable
intervening variable
multiple regression
prediction studies
predictor variable
relationship studies
restriction of range
significance level

DISCUSSION QUESTIONS OR ACTIVITIES

1. A researcher wants to know whether children of parents who are teachers are more likely to become teachers themselves than are children whose parents are not teachers. Discuss how you might conduct this study using a causal-comparative approach. In your discussion, consider how you would define the independent and dependent variables, how you would select the sample, and how you would control for possible extraneous variables.
2. Your school is attempting to develop procedures for predicting which applicants for teaching positions will be the most effective teachers during their first two years of teaching. Based on your study of educational research, you recognize that a correlational prediction study would be one way to examine this issue. Identify possible predictor variables that could be measured for each of the applicants. Also discuss how you might measure their teaching effectiveness. Finally, indicate which predictor variables you believe would have the highest correlation with the criterion variable of teaching effectiveness.
3. Compile a list of variables or events that seem to be correlated but that are unlikely to be causally related (such as the broken bones and weather example in the chapter). If possible, identify variables or events common in

educational settings. Discuss reasons why people are likely to falsely believe that correlated events are causally connected.

SAMPLE CAUSAL-COMPARATIVE STUDY

Beier, S. R., Rosenfeld, S. D., Spitalney, K. C., Zansky, S. M., & Bontempo, A. N. (2000). The potential role of an adult mentor in influencing high-risk behaviors in adolescents. *Archives of Pediatric and Adolescent Medicine, 154*(4), 327–331.

SAMPLE CORRELATIONAL STUDIES

Dimopooulos, D. I., & Pantis, J. D. (2003). Knowledge and attitudes regarding sea turtles in elementary students on Zakynthos, Greece. *Journal of Environmental Education, 34*(3), 30–38.

Fisher, J. L. (1995). *Relationship of intelligence quotients to academic achievement in the elementary grades.* ERIC Document Reproduction Service No. ED 388428. Retrieved April 14, 2005, from ERIC database.

SUGGESTED READINGS

Johnson, B. (2000, April). *It's (beyond) time to drop the term causal-comparative and correlational research in educational research books.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA (ERIC Document Reproduction Service No. ED445010).

CHAPTER ELEVEN

INFERENTIAL STATISTICS

CHAPTER OBJECTIVES

- Differentiate between statistics and population parameters
- Explain the purpose and use of inferential statistics
- Understand the concept of a true difference and chance difference
- Explain the concept of sampling error and its relationship to the generalizability of a study
- Understand the standard error of the mean and be able to construct confidence intervals around the sample mean
- Explain the relationship between tests of significance and the null hypothesis
- Know the concept of statistical significance and how it relates to research
- Understand rejecting and failing to reject the null hypothesis

BEYOND DESCRIPTIVE STATISTICS: INFERENTIAL STATISTICS

Although researchers use descriptive statistics to summarize sample data, in certain types of studies researchers want to draw conclusions about the population from which the sample was drawn. In this case, numbers that they calculate based on data from the sample (e.g., the mean and standard deviation) are called **statistics.** These statistics are used to make estimates about the corresponding values in the population that are called **parameters**. In inferential statistics, the sample values are our best estimate of what the population parameters are likely to be. Correlational studies, causal-comparative studies, and experimental studies may use inferential statistics to draw conclusions about population parameters. As professors of educational research whose classes usually comprise practitioners, we often struggle with how much information we should present to our students about inferential techniques. How much is too much information that is not relevant to the professional lives of practitioners? How much information is needed for the practitioners to understand the research they might read and allow them to become critical consumers of research? We have tried to achieve a balance by presenting the underlying concepts critical to the understanding of inferential statistics as they relate to correlational, causal-comparative, and experimental studies. These concepts are similarly applied across all quantitative studies that use inferential techniques, with slight variation depending on the type of study, the type of data, and the type of statistical tool used. We also present a sample of inferential statistical tests commonly used in educational research.

Underlying Concepts

As we have indicated, when experimental researchers conduct studies, their goals are generally twofold. First, they want to determine whether an experimental treatment causes a difference between the groups being studied. Second, using randomization procedures, they hope to generalize or draw similar conclusions about the population. Therefore, drawing conclusions about the population based on the sample results is an important component of experimental research. In order to draw those conclusions, researchers must use inferential statistics.

Consider the following example presented in chapter 9: A researcher conducts a study to determine the effectiveness of computerized tutoring on the scores of seventh graders on a standardized math achievement test. The researcher for the school district randomly selects 60 seventh graders from a population of 175 seventh graders enrolled in an after-school tutoring program. Thirty students are randomly assigned to the experimental group (individualized computerized tutoring) and 30 students to the control group (no computerized tutoring; these students attend the after-school program in which a teacher helps them with their math homework). Note that if the computerized tutoring proves to be effective, all seventh graders enrolled in the after-school program will be given the opportunity to participate in the computerized tutoring program. Even though the researcher randomly selects the students, she administers a math pretest to determine whether the two groups are equal at the start of the study. Finding the groups to be the same, the study proceeds for 6 months, and the researcher administers a posttest (a standardized math test). She calculates sample statistics and finds that the mean performance of the experimental group exceeds the mean performance of the control group. Although the researcher is happy with the results (e.g., the sample data indicate that the treatment made a difference), she knows that another goal of experimental research is to generalize the findings back to the population. The researcher asks the following questions:

- How likely is it that the difference between the sample means would also be true of the population means? That is, how confident can the researcher be that the difference found between the sample means represents similar differences in the population parameters?
- Are the differences obtained between the two groups due to the treatment or to chance?

The only way to answer this question with 100% certainty would be to conduct the experiment with the entire population. Because most researchers are not likely to conduct the study with the entire population, they use inferential statistics and probability to determine the answer. So, in the case of our computerized tutoring example, the researcher would want to know if similar results would be obtained if the entire population was used in the study. In other words, does the difference found between the computerized tutoring and noncomputerized tutoring group reflect a **true difference**, or is the difference due to chance? The word *true* has technical meaning. A true difference indicates that the results collected on the sample in fact reflect the population parameters; that is, the results would be similar if we did measure the whole population, allowing for generalizability. In addition, a true difference is one from which it is appropriate to conclude that the difference found between the groups is due to treatment (the computerized tutoring) and not merely to chance. How did our researcher try to ensure the generalizability of the study? If you said by using randomization, you are right! However, although randomization is the best way to increase the likelihood that the sample is representative of the population, it does not guarantee that the sample represents the population. You might be asking, "Why not?" Read on and you will find out!

Sampling Error What are the chances that even when using a randomization technique, the researcher in the previous study would end up with a sample that was identical to the entire population of seventh-grade students? If your answer was almost zero, you are right! There is always going to be variation between the sample and the population. This variation that you would expect in any experimental study is called sampling error. **Sampling error** is the expected chance variation that exists between the sample and the population. Keep in mind that as your sample size increases, the sample will be a more accurate representation of the population and the less sampling error you will have.

However, as a researcher, you will always have sampling error unless you do not select a sample and you use the entire population! So how can you end up with a sample that represents the population as accurately as possible? In other words, how can you get a mean from a sample to be an accurate reflection of the mean of an entire population? One complicated and impractical way would be to draw multiple samples from the population and calculate the mean of the means. Sounds confusing, right? Well, consider this. The mean of the means will likely be a good estimate of the population mean. You would expect each sample mean to be somewhat different just by chance. But taken together, calculating the mean of the means would probably get you close to the population mean. The best way to explain this is by giving you actual data. Figure 11.1 shows a distribution of 100 scores, or our population.

Let us say that I randomly select a sample of 15 scores and calculate the mean. Sample 1 includes the following numbers: 87, 56, 99, 76, 67, 89, 56, 34, 97, 67, 71, 66, 89, 99, and 98. If I calculate the mean, there is likely going to be some sampling error, right? Right. Every sample will have some sampling error or variation due to chance. In fact, the mean of sample 1 is 76.73, and the population mean is 68.45; the difference between these two numbers is due to sampling error.

Now let us take two more random samples of 15 scores. Sample 2 is made up of the following: 76, 78, 90, 55, 78, 18, 90, 23, 87, 29, 34, 45, 67, 99, and 88, and the mean is 63.8. The scores for sample 3 include 98, 35, 75, 93, 67, 98, 56, 99, 12, 88, 76, 74, 93, 52, and 87. The mean for sample 3 is 73.53. As you can easily see, there is error associated with each of the random samples. None of them is exactly equal to the population mean.

Now let us take the mean of the samples 1, 2, and 3, or the mean of the means. This mean is approximately 71.35 and is a closer estimate of the population mean.

FIGURE 11.1 A Distribution of 100 Scores

76	23	45	67	89	90	98	87	76	78
34	46	78	90	34	78	23	56	87	98
80	67	98	99	55	67	69	34	67	98
76	56	78	89	23	35	76	87	98	60
45	67	87	98	90	76	56	45	43	65
76	88	99	88	98	87	65	54	78	55
45	76	89	12	34	56	87	89	90	87
67	86	45	45	66	88	99	87	66	55
34	23	67	87	79	75	57	67	54	23
97	29	18	71	73	91	74	93	52	87

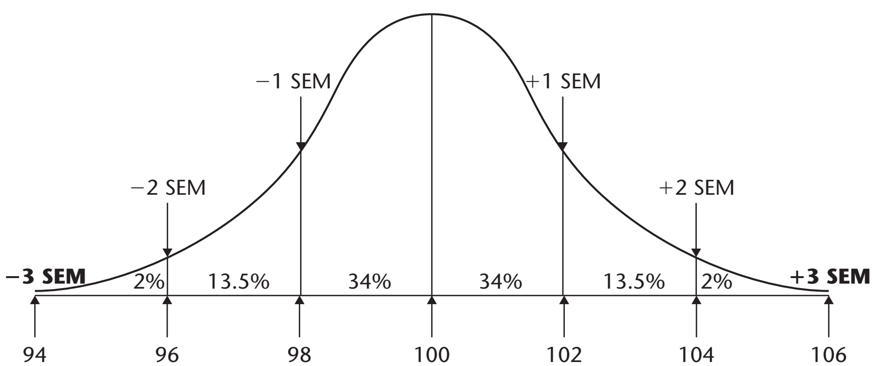
Now, what is the chance that a busy researcher is going to want to select multiple samples and conduct the study several times? You are right if you said never! So what is a researcher to do? Well, there are some interesting characteristics of sample means. First, sample means are thought to be normally distributed. Just what does that mean? Well, if a large number of sample means is selected from a population, the means will form a normal distribution. The distribution of sample means is like any other normal distribution in that it has a mean and a standard deviation. Except in this circumstance, the standard deviation has a special name: the **standard error of the mean (SEM)**. Using the characteristics of the normal curve, the sample mean, sample standard deviation, and the SEM (calculated with a relatively simple formula shown below) helps the researcher to estimate where the population mean is likely to fall. N equals the number of subjects in the sample.

$$\text{STANDARD ERROR OF MEAN} = \frac{\text{STANDARD DEVIATION}}{\sqrt{\text{N}-1}}$$

Estimating the population mean requires that the researcher use probability and the characteristics of the normal curve (Figure 11.2).

Before you throw your hands up in total confusion, let us use some actual data. Let us say that you randomly select a sample from a population; the sample mean is equal to 100 and the standard deviation is equal to 12. Based on our earlier discussion, what do you know about this mean? You know that there is some sampling error associated with this mean. All we can do is estimate where the population mean is likely to fall because we are going to use only one sample. The way we can estimate the population mean is to calculate the SEM. In our example, let us say that we calculate the SEM using the previous formula and find that it is equal to 2. We can then build **confidence intervals** around the sample mean and predict with probability where the population mean is likely to fall. A confidence interval uses the principles of the normal curve and probability to estimate the range of values that the mean might assume. Remember from our discussion in chapter 3 that there are fixed percentages between the mean and each standard deviation under the normal curve. About 68% of the scores fall in the middle of the normal curve between the mean and plus or minus one standard deviation. About 95% fall in the middle of the normal curve between the mean and plus or minus two standard deviations. Finally, about 99% fall in the middle of the normal curve between the mean and plus or minus three standard deviations.

FIGURE 11.2 Normal Curve



So, in this situation, based on a sample mean of 100 and an SEM of 2, we can predict the following:

- With 68% confidence, the population mean will fall between 98 and 102.
- With 95% confidence, the population mean will fall between 96 and 104.
- With 99% confidence, the population mean will fall between 94 and 106.

Note that as a researcher, you would want your SEM to be as small as possible. Why? A small SEM would mean that your confidence intervals would be small, and you would be more likely to know where the mean of the population falls. The key to a small SEM is sample size. Thus, as the size of the sample increases, the size of the SEM decreases. This is perfectly logical! *Finally,* you might say. If the sample size is large, then it must be more representative of the population, and therefore, your sampling error and SEM will be smaller.

The Null Hypothesis, Statistical Significance, and Error All statistical procedures test the null hypothesis, which you may recall from chapter 9 is a statement about the population. That is, in experimental studies, the null hypothesis tests whether the difference found between the sample means is due to chance (sampling error) and not treatment. The chance explanation is the null hypothesis. The goal of the researcher generally is to reject (not support) the null hypothesis and to accept (support) the research hypothesis. In our example, the null hypothesis would be that there would be no difference in achievement between the computerized tutoring group and the noncomputerized tutoring group and that any difference found was due to chance. The researcher wants to reject this null hypothesis and conclude two things: (1) that the treatment made a difference and (2) that the difference also exists in the population (remember the generalizability of the study).

What allows the researcher to make a decision about the null hypothesis? How does the researcher know if the null hypothesis is true (the difference was due to chance or sampling error) or false (the difference was due to treatment)? To be able to make these conclusions, the researcher applies a test of significance to the data collected in the study. **Tests of significance** (to be discussed more at the end of this chapter) are statistical tools that allow the researcher to make decisions about the null hypothesis.

Consider the following. At the end of an experimental study, there is likely to be at least some difference between the experimental and control groups. How does the researcher know if the difference is large enough to conclude that the difference was due to treatment and not simply to chance? To make such a conclusion, the researcher applies a test of significance to the data. Given that the researcher's conclusions will be based on data from the sample, any conclusions drawn must be based on probability. Remember that we did not test the entire population but only the sample. The researcher sets the acceptable probability level at the start of the study, before the data are even collected. In fact, it is written into the research proposal in studies that collect data, or at least it should be! In education, the generally acceptable probability level (p value) is less than .05, or 5 out of 100. What does this mean? It means that the researcher will not conclude that there is a true difference (rejecting the null hypothesis) unless the probability of obtaining the differences by chance is less than 5% out of 100%, or .05. This is translated into **statistical significance**. In other words, the researcher is 95% confident that the difference reflects a true difference, is due to treatment, and therefore represents a statistical significance. At times, the researcher will set the probability level at .01; he or she is willing to tolerate only a 1% out of 100% risk that the difference is due to chance and not treatment. In these cases, the differences between the means must be large. When reading research, how will you know if the study has found statistical significance? Simple. The researcher will indicate in the narrative of the findings section that statistical differences were found and will also report in a table of findings the type of test and its numerical outcome. The numerical outcome will have an asterisk (*) next to it or a p value ($p < .05$) indicating statistical significance.

Let's apply these concepts to a study in which the sample mean of the experimental group is equal to 100 and the mean of the control group is equal to 110. The researcher wants to determine whether the difference between these two sample means is large enough to infer that a similar difference exists in the population. In order to make this decision, the researcher applies a test of significance (for example, a t test, ANOVA, etc.). These tests of significance produce a value or an output (sometimes presented in a results section as a t or F value) that is compared to a table listing critical values for that test. (For one example of this type of table, see HyperStat Online at http://davidmlane.com/hyperstat/t_table.html.) Then using the probability level established prior to the study, the researcher decides whether to reject or fail to reject the null hypothesis stating that no difference exists in the population.

Because these decisions are always based on probability, there is always a chance that an error will be made. To examine these errors, let us look at the choices the researcher has relative to the null hypothesis:

Choice 1: The researcher concludes that the null hypothesis is true. If the researcher concludes that the null hypothesis is true, he or she is saying that any difference between the sample means is due to chance and not the treatment. This means that the researcher is concluding that the difference between the sample means is too small to conclude that the population means differ. Let us say that in the population, the null hypothesis is in fact true. The researcher has made a correct decision and fails to reject the null hypothesis. Yeah!

Choice 2: The researcher concludes that the null hypothesis is false. By concluding that the null is false, the researcher is saying that the difference found is due to treatment and not to chance. Great news for the researcher! The researcher is saying that the difference found between the sample means is large enough to draw the conclusion that a similar difference also exists in the population. In reality, the finding from the sample is also true of the population. The researcher has once again made the correct decision, only this time, he or she will reject the null hypothesis. Right again!

Choice 3: The researcher concludes that the null hypothesis is false, but in reality (i.e., in the population), it is true. Sorry, the researcher has now made a mistake. The researcher incorrectly concludes at the end of the study that the treatment made a difference and therefore rejects a null hypothesis that is actually true.

Choice 4: The researcher concludes that the null hypothesis is true. So, what is being said here? The difference found was due to chance and not treatment. But in this case, the true state of the population suggests that the difference was in fact due to treatment. Once again, a mistake has been made. The researcher fails to reject the null hypothesis and draws the wrong conclusion.

These mistakes have names. In choice 3, the researcher has made a type I error: rejecting a null hypothesis that is really true. In choice 4, the researcher has made a type II error: failing to reject a null that is false. The probability level or significance level ($p < .05$ [or .01]) determines the chance of making a **type I error**. So, if the selected probability level is .05, the researcher has a 5% chance of making a mistake or a type I error. Obviously, if the researcher sets the probability level at .01, there is only a 1% chance of making a type I error. The less chance of being wrong you are willing to tolerate, the greater the difference between the sample means is needed to reject the null hypothesis and say that the difference was due to treatment. What if the researcher was not willing to chance making a type I error and sets the probability level at $p < .0001$? Well, in this case, the researcher would have a very small chance of making a type I error but would increase his or her risk of making a **type II error**. The researcher would hardly ever reject a null hypothesis and may miss an actual significant difference. See Exhibit 11.1 for an illustration of the difference between type I and type II errors.

EXHIBIT 11.1
Type I and Type II Errors

Researcher's Conclusion	True Status of Null Hypothesis: True	True Status of Null Hypothesis: False
True	The researcher is right!	Type II error
False	Type 1 error	Right again!

Source: Adapted from Gay and Airasian (2003).

Understanding the conceptual framework of inferential statistics should help you to read and understand some of what is written in the results sections of research studies. However, you should know that we have simply provided you with an overview of the concepts that are critical to a basic understanding of inferential statistics. There are many more complicated concepts that you will learn about if you continue your studies beyond the master's degree level.

STEPS IN ANALYZING DATA USING INFERENTIAL TESTS

With a basic understanding of hypothesis testing, let us turn our attention to the process used by researchers who conduct quantitative studies in which they use inferential techniques. As is we hope is clear, a basic component of the inferential process is to test the null hypothesis and make a decision about its veracity. There are systematic steps that a researcher takes to engage in this investigation. In fact, many of these steps are taken before the analysis stage of the research. They are as follows:

Step 1: Review the Null Hypothesis

Having stated the research hypothesis (or alternate hypothesis) at the conclusion of the review of the literature, the researcher reviews the null hypothesis. Remember that the null hypothesis states that there is no real difference or relationship between the groups or variables and that any difference found was due to chance. For example, the research or alternate hypothesis is as follows:

> *The computerized tutoring group will perform better on a standardized achievement test than the noncomputerized tutoring group.*

The corresponding null hypothesis would be that

> *There is no difference in the performance on a standardized test between the computerized tutoring and the noncomputerized tutoring group.*

Step 2: Decide on Probability Level

Before the data collection and after the researcher clearly states the null hypothesis, the researcher sets the significance level or the probability level (p value). That is, what criteria will you use to reject the null hypothesis and conclude that the difference you found was due to treatment and not to chance? The researcher usually decides to set the probability level at either .01 or .05. Remember what this means? If you set the probability level at .01, 1 out of 100 times, the difference is due to chance, and if you set it at .05, the risk is 5 out of 100 times that the difference will be due to chance. Most educational researchers set the probability at .05, preferring not to miss a true difference that might exist.

Step 3: Select the Statistical Tool

Although we are defining the selection of the statistical tool as step 3, in reality, many researchers know what test they are going to use even before they begin their research. Many researchers simply prefer one statistical tool over another. However, if they have not made a decision, here is where they must select the statistical test. Many statistical tests are used in educational research, and some of the most commonly used ones are listed in Table 11.1.

The researcher decides on the statistical test to use in part based on the type of data, type of hypothesis, and the number and type of variables in the study. One of the first decisions related to choosing the statistical technique is to decide whether to select a **parametric** or a **nonparametric** test. Both parametric and nonparametric tests allow you to "draw inferences about population based on samples" (Cronk, 2008, p. 53). Parametric tests are generally preferred because they are believed to be more powerful tests. A *powerful test* is one that is more likely to allow the researcher to reject a false null hypothesis. However, to use a parametric test, the following assumptions about the data must exist:

TABLE 11.1 Commonly Used Inferential Statistical Tests

Statistical Test	Type of Data	Number and Type of Variables	How Used and Interpretation
t Test	Interval	1 independent; 1 dependent	Used to test the difference between two group means; a significant *t* value shows that a true difference exists between the group means
Analysis of variance	Interval	1 or more independent; 1 dependent	Used to test the difference between the means of two or more groups; a significant *F* ratio shows that a true difference exists between the group means
Analysis of covariance	Interval	1 or more independent; 1 dependent	Used to test the difference between the means of two or more groups after the influence of an extraneous variable (covariate) has been statistically removed; a significant *F* ratio shows that a true difference exists between the group means
Chi-square	Nominal (frequency counts or Percentages)	2 or more frequency counts or percentages	Used to test whether the observed frequencies (the frequency counts or percentages from the data) show a true difference from the frequencies expected if all categories were equal
Pearson product-moment correlation	Interval percentages	2 variables that are both measured	Used to test whether the relationship between two variables is greater than would be expected due to chance; a significant *r* shows that a true relationship exists
Multiple regression analysis	Interval	2 or more independent; 2 or more dependent	Used to see if the independent variable predicts changes in the dependent variable when other variables are held constant; a significant *R* value means that the independent variable can predict differences in the dependent variable

- The data must be interval or ratio.
- The subjects must be independent of one another (i.e., randomly selected).
- The variable being measured must be normally distributed in the population, or at least the type of distribution must be known.
- The variability of the dependent variable in the two population groups must be the same. This is a statistic called the **variance** (which as noted in chapter 3 is the standard deviation squared).

Some violation of these assumptions is acceptable, with the exception of independence. This means that even if all of these assumptions are not met, parametric tests may still be used. Nonparametric tests are used primarily when the data are nominal or ordinal or whenever the assumptions for parametric tests are gravely violated.

Step 4: Calculate the Results of the Statistical Test and Make Decision About Rejecting the Null Hypothesis

Following the collection of the data, the researcher then calculates, typically with computer software programs, the selected statistical test. The statistical test will produce a value. For example, let us say that our researcher who is examining the difference between computerized and noncomputerized tutoring decides to compute a *t* test at the .05 level of significance on the achievement data. The *t* test will be used to determine if the researcher should reject the null hypothesis. The *t* test will produce a *t* value, and the researcher will compare that value with those of a table of *t* values (or a computer printout of these values).

Each statistical test has associated with it degrees of freedom (*df*). The concept of degrees of freedom is beyond the scope of this book, but typically they are equal to the number of groups minus one. All practitioners really need to know is that to compare the obtained value with the tabled value, you must know the degrees of freedom. Even though this will likely be done for you by the computer program, we describe here how it was done in the days before we used computers to analyze data.

Let us say that the calculated *t* value in our study is 8.4. At the .05 level of significance and with 1 *df*, the table value is equal to 12.71 (trust us, we looked this up). Because our value is 8.4, it is less than the table value. Therefore, our *t* value is not large enough for us to reject the null hypothesis. What does this mean? It means that the difference between the means was not large enough for us to reject the null hypothesis. The difference we found was due to chance and not to the treatment, in this case, the computerized tutoring.

DESIGNS WITH MORE THAN ONE INDEPENDENT OR DEPENDENT VARIABLE

Some research designs include more than one independent or dependent variable. Under those circumstances, the analyses become a bit more complicated. Although we do not discuss all of these designs, we illustrate one design with more than one independent variable, referred to as a **factorial design**.

There are times when researchers want to examine more than one independent variable (as was the case with the example of computer use and class size) or at least one independent variable and a **control variable**. A control variable is a variable that the researcher has identified as one that could confound the outcome of the research study. A good example of a control variable can be seen in the study on computer use and gender. As a good researcher, you conduct a literature review investigating the existing research on computer use as an instructional tool to improve the math performance of elementary school students. You discover that a recurring theme in the literature is that boys and girls react differently to computer use and that this reaction can influence math achievement, the dependent variable in the study. So, perhaps either boys or girls will do better when exposed to computer use (the experimental treatment). To investigate not only computer use but the potential **interaction** of computer use and gender, the researcher could set up a 2×2 factorial design. A 2×2 factorial design has two factors (or independent variables) and two levels. The method of instruction is one factor, and the two levels are computer use versus no computer use. Gender is the other factor, with boys and girls representing the two levels. Thus, a 2×2 design includes four cells, as depicted in Exhibit 11.2.

EXHIBIT 11.2
A 2 × 2 Factorial Design

	Boys	Girls
No computer use	Cell 1; boys who get no computer use	Cell 2; girls who get no computer use
Computer use	Cell 3; boys who get computer use	Cell 4; girls who get computer use

To explain further the concept of interaction, Exhibit 11.3 includes hypothetical outcomes for the dependent variable math achievement.

Our hypothetical math achievement test consisted of a 100-point exam. We have reported the cell means for each group. Researchers might examine the **main effects** (the primary or main variables or factors in the study). The primary variables in the study were method of instruction and gender. To examine these main effects, the researcher would collapse the cells and obtain an overall mean. So, for the no-computer-use group, the mean or average performance was 77.5 (cell 1 plus cell 2 divided by 2). The mean performance for the computer-use group was 87.0 (cell 3 plus cell 4 divided by 2). Based on these means, the computer-use group outperformed the no-computer-use group. In addition, when examining gender, the mean performance for the boys was 74.0 (cell 1 plus cell 3 divided by 2) and 90.5 (cell 2 plus cell 4 divided by 2) for the girls. This suggests that the girls outperformed the boys. However, was there an interaction between gender and method of instruction? The main effect of computer use suggests that computer use was better than no computer use, but was this true for both boys and girls? Examine the cells. In reality, the boys did better under the computer-use

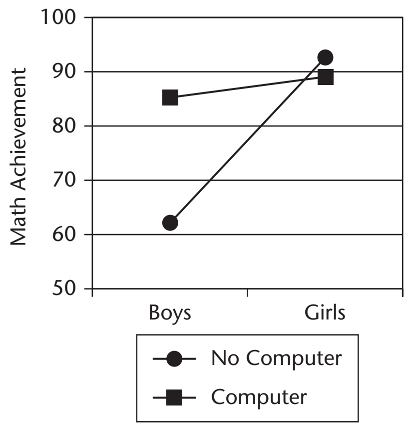
EXHIBIT 11.3
A 2 × 2 Design Showing Mean Scores of Boys and Girls on a Math Achievement Test

	Boys	Girls	Means for Computer-Use and No-Computer-Use Groups
No computer use	Cell 1; mean = 63.0	Cell 2; mean = 92.0	77.5
Computer use	Cell 3; mean = 85.0	Cell 4; mean = 89.0	87.0
Means for all boys and all girls	74.0	90.5	

These means show the main effect of computer use or no computer use

These means show the main effect of gender on math achievement

FIGURE 11.3 Graph Showing Interaction of Gender and Computer Use



condition, and the girls did equally well under both conditions. Such a result might suggest that computer use is an effective way to improve math achievement in boys, but it has little effect on the math achievement of girls. Many researchers choose to diagram interaction effects using a graph. Interaction effects are depicted as nonparallel lines, as shown in the graph in Figure 11.3.

Although it can be difficult to understand all the numbers involved in complex inferential tests, such as a factorial design, remember that complex analyses allow us to examine multiple variables. By examining multiple variables, we get closer to accurately describing the complexities of real life. No one thinks you can accurately describe a classroom using only one or two variables! So, inferential statistics are simply another tool for increasing our understanding of the many factors that influence educational processes and outcomes.

SUMMARY

Although researchers use descriptive statistics to analyze data, they must use a higher level of analysis when making inferences about the population from which the sample was drawn. Tests used to make inferences about a population based on a sample are referred to as inferential statistics. Correlational studies, causal-comparative studies, and experimental studies employ the use of inferential statistics to draw conclusions.

In experimental research, the researcher wants to know whether the experimental treatment caused a difference between the two groups and hopes to make generalizations of the findings back to a wider population by using randomization procedures. Even though an experimental researcher goes to the trouble of using randomized procedures to ensure that the sample is representative of the larger population from which it was drawn, sampling error can occur. Sampling error is not the fault of the researcher but is merely a reality of randomization.

When testing a hypothesis, the researcher's goal is to reject the null hypothesis and accept the research hypothesis. In making these conclusions, a researcher applies a test of significance. This test uses statistical tools that allow the researcher to make decisions about the null hypothesis. Commonly used inferential statistical tests include *t* test, analysis of variance, analysis of covariance, chi-square, Pearson product moment correlation, and multiple regression analysis.

KEY CONCEPTS

control variable	parameters	statistics
factorial design	parametric or nonparametric	test of significance
interaction	sampling error	true difference
main effects	standard error of the mean	variance

DISCUSSION QUESTIONS OR ACTIVITIES

1. Pick an experimental study that you have read. See if you can find examples of concepts covered in this chapter in the results sections.
2. In quantitative research, statistical tests are often the most complicated parts of a research study. Discuss why these tests are needed and how they relate to the concepts of sampling error and generalizability.

SUGGESTED READINGS

Almer, E. C. (2000). *Statistical tricks and traps: An illustrated guide to the misuses of statistics*. Los Angeles: Pyrczak Publishing.

Best, J. (2001). *Damned lies and statistics: Untangling numbers from the media, politicians, and activists*. Berkeley: University of California Press.

Creighton, T. B. (2001). *Schools and data: The educator's guide for using data to improve decision making*. Thousand Oaks, CA: Corwin.

CHAPTER TWELVE

ACTION RESEARCH

CHAPTER OBJECTIVES

- Define action research
- Discuss how action research differs from applied research
- Discuss the historical and philosophical roots of action research
- Identify the advantages of action research for educational practitioners and how it contributes to professional knowledge about education
- List the key characteristics of action research
- Discuss the different types of action research and those individuals who have played a major role in the development of this area of research
- Identify the steps involved in designing an action research study and illustrate these steps using a self-selected topic
- Identify types of data that might be collected for an action research study and discuss how to select and compare methods for a study
- Understand the criteria for evaluating action research

RESEARCH VIGNETTE

Yolanda is a middle school math teacher who is taking a professional development workshop on action research. She decides to focus her action research study on teaching students to add and subtract positive and negative numbers, an area that her students have trouble understanding. She has just completed teaching a class using the "rule" method, in which she presented the rules for adding and subtracting positive and negative numbers on the board, illustrated the rules with example problems, and then had her students complete sample problems for which they had to apply the rules. Following her presentation and class activity, Yolanda gave a quiz, and 15 out of 20 students failed. So Yolanda's initial research question became, "How can I improve my students' understanding of positive and negative numbers?"

She reflected on her classroom practice with other teachers in the workshop and realized that students were inattentive and unresponsive during the lesson. She realized that her didactic instructional approach gave her students little opportunity to think through the rules on their own. She decided to examine some alternative ways of reteaching the lesson by conducting a review of literature on approaches to math instruction that emphasize active student involvement. One article described small groups of students learning about addition and subtraction by playing a card game that represented negative numbers as red cards and positive numbers as black cards. The game was structured so that students had to figure out the mathematical rules in order to do well in the game. Yolanda decided to try the card game strategy with her class and designed an action research study to see whether it improved her students' understanding.

Yolanda introduced the card game during her next class and collected data by observing and videotaping the student groups as they played the card game. She then gave students a quiz with problems similar to their first quiz. She interviewed some of the students with questions about the activity: what they liked, what they didn't like, and how the activity changed their thinking when adding and subtracting positive and negative numbers. She analyzed her data and found that all but three of her students passed the second quiz and got 85% to 90% or more of the questions correct. Interview data indicated that the students liked the gamelike nature of the task and thought it helped them figure out why they got some items on the first quiz wrong. Yolanda's examination of the video revealed evidence that supported the interview data. Several students said aloud as they were playing the card game, "Oh, this is why I got it wrong on the test," or "Oh, I get it now, I really get it, it is so easy—here let me show you." In addition, Yolanda noticed that as the level of the game increased in difficulty, students slowed down when passing cards to other students and were carefully thinking about what they were doing. It was evident that they were thinking more deeply than they had when Yolanda used the rule method. Yolanda continued to discuss her interpretations of the data with the teachers in her workshop throughout the study and showed them portions of the videotapes and interview data. Their discussions helped her to identify some ideas that she had missed and provided support for other interpretations of the data.

Yolanda concluded that the strategy was effective in teaching students about positive and negative numbers. She planned to continue the strategy with future classes and collect data to see whether it was effective. However, she worried that she might not have time to teach all math concepts using approaches that involve active student learning. She decided to continue to explore strategies and talk with other teachers about how they cover their curriculum while using strategies that involve active student learning and discovery.

UNDERSTANDING ACTION RESEARCH

Action research is a type of research that has been used in many disciplines, including education. **Action research**, as its name implies, within education is a type of research that aims to enact immediate changes in an educational setting. It has the potential to produce change quickly because the research is carried out by educators in their own work settings. Action research borrows techniques and ideas from all other types of research but differs from the purely quantitative or qualitative approaches in that its orientation combines professional practice, research, and reflection on one's own educational practices (Arhar, Holly, & Kasten, 2001). This means that it simultaneously serves to enhance the professional skills of educators, advance our knowledge, and improve educational processes and outcomes. Educators involved in action research develop both personal knowledge and sensitivity about their practices and contribute to the professional knowledge of their field. Reason and Bradbury (2008) have described action research as a "living inquiry that links practice and ideas" (p. 1) and creates collaborative learning communities to facilitate ongoing renewal of education. Through action research, practicing educators study themselves and their learning communities as they try to change and improve educational processes and outcomes. Both quantitative and qualitative applied researchers aim to improve practice through the knowledge generated by first conducting studies and then reporting their results in journals or at conferences. However, action researchers aim to transform their educational practices as they study them. One metaphor used to describe action research is that educators are "making the road while walking it" (Reason & Bradbury, 2008, p. 16). Yolanda illustrates this as she studies her own teaching and her students' learning at the same time as she is trying to improve it.

Some researchers view action research as using primarily qualitative methods, and there are strong similarities between qualitative and action research. In both qualitative and action research, researchers continually examine their own assumptions and biases because they are active participants in the setting that they study. Many action researchers, like Yolanda, collect qualitative data through interviews and observations. However, there are important differences between qualitative and action research. Qualitative researchers often aim primarily to accurately represent and describe the settings that they study. They work hard to blend into the settings that they study and to avoid changing the "natural" interactions and outcomes in those settings. By contrast, the major purpose of conducting action research in education is to change and improve educational settings and outcomes. Action researchers do not have to blend in because they are already active practitioners in their educational settings. In addition, many action researchers collect both quantitative and qualitative data, taking the pragmatic viewpoint that any data that helps them identify "what works" is good information.

Regardless of whether qualitative or quantitative data are collected, all forms of action research have as their major goal improving education. Mills (2000) offers a general definition of action research as "any systematic inquiry conducted by teacher researchers, principals, school counselors or other stakeholders in the teaching/learning environment, to gather information about the ways that their particular schools operate, how they teach, and how well students learn" (p. 6).

Historical and Philosophical Roots of Action Research

Different forms of action research have originated in many different fields besides education, including organizational development, social work, community organizing, political science (especially civil rights), feminism, urban and economic development, and public health promotion. Due to its complex history and interdisciplinary nature, there have been many different terms used to refer to the different forms of action research. Common terms used for action research in education are *teacher research* (Cochran-Smith & Lytle, 1993), *practitioner research* (Anderson, Herr, & Nihlen, 1994), *participatory action research* (McTaggart, 1997), and *emancipatory praxis* (a term emphasizing the need for practices that liberate persons and promote optimal growth and development) (Lewin, 1948). Other terms are sometimes used to highlight certain aspects of action research, such as its scientific nature (*action science*), its focus on seeking knowledge through reflection with others (*collaborative inquiry* or *participatory inquiry*), and its roots in humanism and emphasis on positive growth (*appreciative inquiry*). In this chapter we focus on the development of action research in education.

The roots of contemporary action research in education are often traced back to John Dewey (1938), an educational philosopher who believed that knowledge and action should not be separated. Dewey emphasized that teachers must critically reflect on the consequences of their actions and subject their ideas and practices to critical testing (cited in Arhar et al., 2001). Dewey believed that by applying scientific and problem-solving methods, educators would model the scholarly skills and socially responsible ideals that they should be teaching their students. Kurt Lewin (1948), a social psychologist, shared Dewey's vision of educators as researcher-scholar-practitioners and viewed action research as a way to challenge established educational practices. Lewin believed that both critical reflection and theory should go hand in hand with practice. He argued that action without reflection and understanding was blind, but theory without action was meaningless (cited in Reason & Bradbury, 2008, p. 4). Both Lewin and Lawrence Stenhouse, the founder of the Center for Applied Research in Education, believed that traditional basic and applied research approaches were too slow and disconnected from the everyday world of education to really improve practice. They encouraged educators to become involved in research as a way to improve their own practices. Stenhouse argued that "research would liberate teachers from the dictates of professional researchers" and said that "researchers [should] justify themselves to practitioners, not practitioners to researchers" (cited in Cochran-Smith & Lytle, 1993, p. 8). Action research first became popular in educational circles during the 1940s and 1950s but fell into disfavor during the late 1950s and early 1960s in the United States as research became associated primarily with experimental approaches (Gall, Gall, & Borg, 2005). British and Australian educators refined action research theory and practices in the 1960s, and it has enjoyed a resurgence in U.S. education over the past twenty years. Educators whose ideas fueled the development of action research are listed in Table 12.1.

Not all action researchers are out to change the system, but all share a belief that practitioners have the knowledge, power, and skills to study and improve their own practices and the lives of those with whom they work. Underlying all action research is the assumption that practitioners are capable of independent action and systematic inquiry into their own educational practices. Furthermore, action research is based on the assumption that as insiders (persons who are already working in educational settings), practitioners have valuable knowledge that needs to form the basis for making decisions about schools. Instead of viewing research as the domain of outsiders, such as professionals in colleges and universities, action researchers emphasize that those involved in the day-to-day practice are the experts who should be involved in study schools and educational processes.

TABLE 12.1 Action Research Pioneers

Educator/Researcher	Contribution to Educational Research and Action Research
John Dewey (1859–1952) Educational philosopher	Emphasized the links between knowledge and action and theory and practice. Promoted teachers' professional responsibility to critically reflect on and test their ideas and practices.
Maria Montessori (1870–1952) Early childhood educator	Worked with teachers and students to conduct naturalistic observations of classrooms. Trained teachers to use observations to understand how students respond to their actions.
Kurt Lewin (1890–1947) Social psychologist	Proposed the term *action research* to describe research by educational researcher-scholar-practitioners who wished to improve their practices. Stressed the need for critical reflection on one's biases and saw action research as a way to enact immediate changes in educational practices and systems.
Stephen Corey (1904–1984) Curriculum theorist and director of Horace-Mann-Lincoln Institute of School Experimentation	Challenged the top-down model of curriculum development and created a model of curriculum development for teachers using action research. Emphasized examining power differentials and promoting critical social consciousness.
Donald Schön (1931–1997)	Advanced the concept of educators as "reflective practitioners."
Paolo Freire (1921–1997)	Brazilian educator developed a model of action research—*emancipatory praxis*—emphasizing democratic collaboration with persons in learning communities to define problems and seek solutions.
Lawrence Stenhouse (1926–1982) Director of the Humanities Project	Founded the Center for Applied Research in Education. Challenged the idea that research should be conducted by outsider specialists. Emphasized that research by practitioners would improve both practice and research.
John Elliott (1938–) Director of the Ford Teaching Project	Founded the Classroom Action Research Network and created a model for action research including cycles of action and research.

Sources: Arhar et al. (2001); Schön, 1987.

Thus, through action research, practitioners can both help themselves and help to advance knowledge in the field of education. The potential benefits to the practitioner and to the field of education are summarized in Table 12.2.

Can you guess which of the philosophical frameworks discussed in chapter 1 could be applied to action research? Trick question! It can take either the advocacy-liberatory framework or the pragmatic framework as its philosophical basis. Its major goal is to find ways to change the lives of everyone involved in education for the better. And information gathered through action research empowers educators to change educational institutions and question established practices. Educators who take an advocacy-liberatory framework often use action research to seek new policies, programs, or resources to assist groups who are denied power within educational systems. These researchers might study the problems of groups, such as students who are retained, students with disabilities, parents who speak English as a second language, or nontenured teachers. In all cases the researcher would work collaboratively with persons in the groups to identify possible solutions. This type of action research is often referred to as **critical action research** or **critical pedagogy.**

Action research based on a pragmatic framework involves looking at issues or problems in one's own classroom, school, or educational setting to see how practice can be improved without concern for overarching theories or broad social issues. Yolanda in our opening vignette is an example of an educator who is just beginning to apply action research to her own classroom and who initially does

TABLE 12.2 Benefits of Action Research to Practitioners and the Field of Education

Establishes a scientific framework for improving educational practices and outcomes
Promotes collaborative learning communities and decreases isolation of practitioners
Facilitates creation of better communication and collaboration between educators and persons in surrounding communities
Increases respect for professional knowledge and skills of practitioners and enhances professional development of educators
Decentralizes control of knowledge and promotes dialogue and communication between practitioners and researchers in colleges and universities
Ensures immediate changes in educational practices
Provides evidence to support or challenge entrenched educational policies and practices
Facilitates development of practice-based theories and theory-based practices
Contributes to better research that addresses real-life problems
Enables practitioners to model scholarly skills and socially responsible ideals for their students

not look much beyond her classroom door. This type of action research is often referred to as *practical action research*. Examples of both critical action research and practical action research are presented in the next section of the chapter.

Types of Action Research

One broad classification of action research uses the two types mentioned previously: critical action research and practical action. The goal of critical action research is liberation through knowledge gathering. Critical action researchers believe that all research should be socially responsible, aimed at enhancing the lives of all persons but especially those who are marginalized or who lack the power to improve their own lives. Critical action researchers analyze the larger cultural and social systems in which educational problems and situations are embedded. They often address issues such as sexism in schools, racism or prejudice embedded in an educational system, physical and emotional barriers to persons with disabilities, inequities in funding schools, or the isolation experienced by persons who are gay or lesbian.

Critical action researchers believe that just as education requires a respectful dialogue between teacher and student, research must be democratic and equitable. All persons involved in the research have valuable knowledge and skills to be used and valued. Research should involve extensive collaboration with members of the community, who are the only persons who have the right to specify what changes would enhance their lives. Finally, critical action researchers believe that action should be informed and linked to values held by the persons whose lives the research examines.

Paulo Freire (1970), a Brazilian educator, developed many of the earliest ideas about critical action research based on his work on literacy with indigenous peoples in Brazil. He proposed that researchers work collaboratively with persons in a community to identify problems that arise from real-life situations and seek solutions to these problems by questioning and reflecting together. Next we present an example of critical action research.

Bruce King (1990) was a biology teacher whose students raised questions about a toxic dump that was less than 2 miles from their school. The dump was labeled hazardous by the Environmental Protection Agency but was not scheduled for cleanup. King encouraged his students to ask questions and pursue research about the issue. In the process, they learned about how societal and political considerations might conflict with ethical decision making and identified strategies that could be used to address the problems posed by the dump. In the process, together the students and teacher created a curriculum for exploring issues relevant to science and society.

Practical action research aims to improve the lives of people involved in education, but the research typically has a more practical orientation. It is based in everyday practice and typically focuses on making small changes at a local level. Although practical action researchers often are not as advocacy oriented as critical action researchers when they begin their research, they may become more politicized as they investigate the problems in their educational setting. For example, a practical action researcher might begin to explore the problems experienced by students who are receiving special education services. If the research reveals that students of Latino origin are unfairly classified as needing special education services at the school, the practical action researcher might decide to take a critical action stance by investigating how these decisions are made. Both types of action research involve *action*, and sometimes it is unpredictable where those actions will lead. Both practical and critical action researchers share a commitment to involving participants in the research as equal partners, as can be seen in the following example of practical action research.

Jordan and Hendricks (2002) conducted a practical action research study aimed at increasing student engagement in literacy learning in their classrooms. Their initial research question was, "How does student choice in learning activities and assessment affect student engagement?" In class discussions, they solicited student input on assignments and set up student contracts allowing them to choose their own learning tools from activities based on seven of Howard Gardner's multiple intelligences. For example, students could write a book report, present oral discussions of the book, or do dramatic reenactments of the book. Cooperative learning groups were formed based on students' preferred learning styles. Groups completed in-class assignments on vocabulary building, did peer-to-peer reading, and discussed the book. Jordan and Hendricks's methods of data collection included student journals, interviews with students, work samples, and an attitude survey of the students. Their data showed that the students liked having input but were uncomfortable with "too much freedom." Their grades improved, but some did not understand the contracts. Jordan and Hendricks were unsure how much the specific book used in the assignments might have contributed to the increase in engagement. They concluded that student engagement had improved but that they needed to continue the research with a new book and refine their assignments and contracts.

Characteristics of Action Research

So far we have stressed how varied action research is. However, the following characteristics are common to all types of action research.

1. *It is conducted in the practitioner-researcher's own educational setting, and the practitioner takes an active part in the research*. The setting can be a classroom, school,

district, or community program; the situated nature of the research enables the research to draw on insider perspectives. It also ensures that the research is based in the reality of everyday educational practices.

2. *It involves critical reflection and systematic planning and procedures throughout all phases of the study.* Action research builds on Schön's concept of educators as "reflective practitioners" who continually examine and question the effectiveness of their practices (Schön, 1987). Reflection is involved at every stage in an action research study, including the initial posing of a question, examination of one's own assumptions and biases, selection of strategies or interventions, and collection and analysis of data. Reflections are processed systematically, using tools such as journals or discussions with critical partners or colleagues.

In addition to reflection, action research involves the systematic collection and analysis of data to facilitate understanding and provide evidence to support or challenge the effectiveness of one's practices. Being *systematic* means that procedures are clearly specified and tools are designed to ensure consistency in data collection. McNiff (1988) says that action researchers must follow a system of "disciplined inquiry" so that any conclusions can be checked against the data to determine whether they are fair and accurate.

In planning and designing studies, action researchers may use techniques from different types of research. Some action research involves pre- and posttests, as in experimental research. Other studies might utilize brief surveys or qualitative interviews. In addition, action research follows principles of good instrument design and considers issues related to validity and credibility. However, all tools and procedures must be appropriate to the educational setting.

3. *It involves collaboration with other educators and persons involved in the educational process.* Collaboration is vital to deep and honest reflection. As action researchers work with others, they engage in critical discussion of their ideas and practices. This reduces the likelihood that personal biases will distort the findings and ensures that multiple perspectives will be considered.

Collaboration promotes a democratic attitude toward research by inviting all participants in a learning community to play an active role in research. Unlike more applied research, action research emphasizes doing research "with" others instead of "on" others. The setting in an action research study can be a single classroom, office, or program. Many educators who are new to action research begin by reflecting on just their own educational settings. They might gradually involve other practitioners at the school, including other teachers, school psychologists, speech therapists, counselors, staff, or the school principal. Students, parents, and community members are frequently involved in action research to assure that all perspectives are heard.

Because many of the measures used in action research are qualitative, this representation of multiple perspectives enhances the credibility of the research.

It also makes the job of balancing the roles of being both a practitioner and a researcher more manageable. As action research becomes embedded in a school's culture, activities can be extended schoolwide to involve multiple classrooms, offices, or programs. Community involvement might be extended to include community boards or links to community-based organizations. Some action research involves whole school districts. As the level of action research is expanded, collaboration becomes more complex but remains essential. Oftentimes school- or districtwide action research includes collaboration with university-based researchers. Researchers at colleges might assist in developing measurement instruments or in collecting and analyzing data. However, the research is characterized by a mutual respect for the expertise that each person brings to the process.

4. *It focuses on taking action to change and improve educational practices.* As mentioned earlier, all action research at some point involves *action*. This action might be as simple as changing an assignment for the next school year, or it might involve a rethinking of how students are graded. At the school or district level, changes in policies for retaining students or assigning students to special education might result from an action research project. One of our students used her action research project to find out what students thought about parental involvement in the classroom. Based on her findings, she changed some of her roles for parents within the classroom.

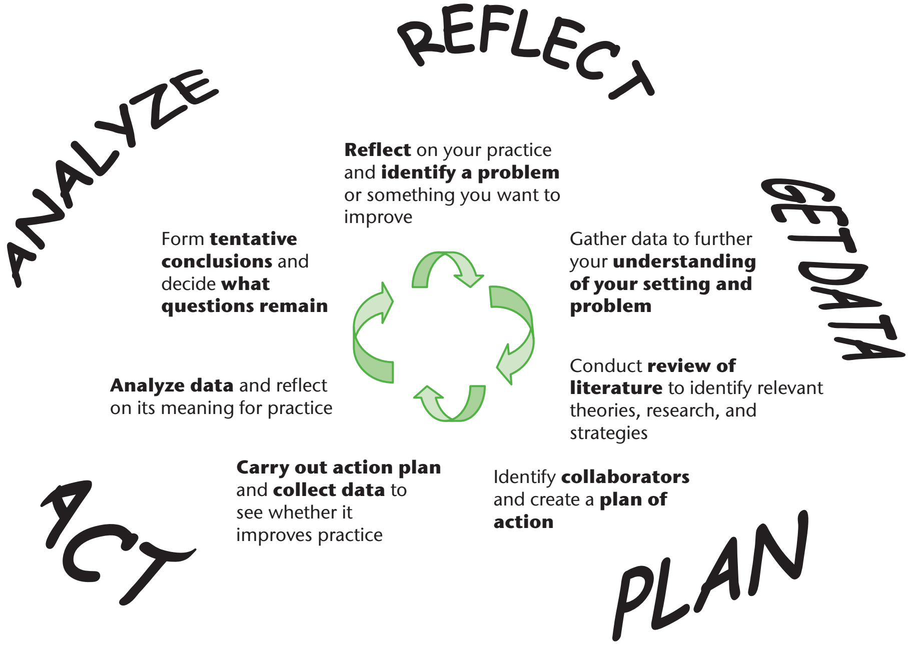
5. *It is ongoing and includes several waves of data collection, reflection, and action.* Because action researchers are educators who deal with problems in their everyday practices, research continues beyond the initial cycle of data collection and analysis.

All good research builds knowledge incrementally. In action research, lessons learned in the initial wave of data collection may lead to new questions, refinements in practice, or identification of new problems. Action researchers are uniquely positioned to continue to collect data and extend their research into these new areas. Therefore, action research typically includes several waves of data collection, reflection on the data, and trying out actions to improve one's practice. The ongoing nature of action research is also consistent with the view that professional educators should continually reexamine and renew their practices.

STEPS IN CONDUCTING ACTION RESEARCH

Action research is fluid and flexible, so it is inappropriate to prescribe a set sequence of steps to follow. One simple way to conceptualize the processes involved in action research is displayed in the outer headings in Figure 12.1. This diagram shows that action research involves five basic processes—reflect, get data, plan, act, and analyze—that are repeated over time. Some theorists describe action research as a spiral that can repeat steps at any point in the process.

FIGURE 12.1 Action Research Process



Although the headings in Figure 12.1 capture the overall processes of action research, they are somewhat oversimplified. The following seven steps describe a more detailed sequence for action research and provide guidance for students setting up their first action research projects. However, remember that as you implement your project, there may be movement back and forth between the steps.

Step 1: Reflect and Identify a Problem

The first step involves reflecting on your practice and identifying a problem or something you want to change. This can be something unusual you have observed in your classroom that piques your interest or mystifies you or something you want to understand in more depth. A teacher might ask, "How can I increase the amount of writing my students do outside of the classroom?" A school psychologist might ask, "How can I improve assessment of learning problems in a child who is bilingual and has difficulties in reading?" An administrator might ask, "What do my teachers believe are the most important resources that our school has or should obtain?" Any question is okay as long as you have access to the setting to collect data and have an ongoing role within it. As with other types of research, it is helpful to begin by writing a research question describing the problem that you want to address.

Action researchers spend a good deal of time reflecting and thinking about their own practice. This is a characteristic of all good educators. However, action researchers often record these reflections in journals or share them with a collaborative group. Arhar et al. (2001) recommend that the initial reflections should be deeply self-critical and should aim to develop as complete a picture as possible of the problem or issue. They suggest that the action researcher consider the following questions:

- What are my feelings about this situation or problem?
- What are my motives and needs?
- What are my assumptions or biases and how might I be wrong? What are the different values of different people in this setting?
- What cultural practices or beliefs affect this problem or situation?
- How well do my practices match up with my values? Do I really practice what I preach or believe? What inconsistencies are there?
- How can I visualize or describe the situation or problem? Would others see it differently?

The practitioner can engage in reflection by himself or herself or through discussion with a "critical colleague." Exhibit 12.1 presents an exchange between two colleagues discussing a student who had appeared to be a daydreamer much of the time. The student, Chuck, made a cross-disciplinary connection that helped many class members understand a math concept.

Reflection continues throughout the action research study.

Step 2: Gather Data

The next step is gathering data to further your understanding of your setting or problem. This might include examination of archival data and writing entries into a reflective journal. One might also conduct observations of their setting or interviews with participants or collaborators. The purpose of this initial data gathering is to more precisely understand the nature of the problem.

Step 3: Review the Literature

In step 3 you conduct a review of literature to identify relevant theories, research and strategies. This usually means reading research (including basic, applied, and action research), theoretical writings, and reflections (including experiential reports) on your topic or problem by other educators. Although action research acknowledges that each setting is unique, researchers seek out information to help them understand what might be going on in their setting. By reviewing published literature, practitioners often find ways to reframe their explanations or discover practices that might work in their own settings. Due to the cyclic and ongoing nature of action research, the review of literature often continues throughout the process of collecting and analyzing the data. By examining their own educational problems or issues in relation to other research or educational theories, action researchers also build upon existing *professional* knowledge. McTaggart (1997) describes action research as a process of "theorizing practice" by developing explanations for why certain practices work and others do not. These mini-theories developed by practitioners might provide support for existing educational theories or challenge them by pointing out how things might work differently for specific students in some classrooms or schools. By reading the literature and later publishing their own findings, action researchers add to professional knowledge.

Step 4: Identify Collaborators and Create a Plan of Action

Collaborators include persons who can help you in your reflections and data collection and analysis. They might be colleagues at your school or site or persons at other schools or colleges who share similar interests. Action researchers frequently collaborate with students, parents, community members, or other persons with

EXHIBIT 12.1
Sample of Two Educators Using Critical Reflection to Define a Research Problem

Chuck's teacher: Here Chuck is, seemingly oblivious to the class, while I am straining to explain a math concept in a way that the children can understand—I try to snap him out of his biological excursion, and instead he leads the class into understanding! How do you figure?

Critical colleague: Yes, strange. You seemed pretty frustrated all right. Did you think he was daydreaming?

Chuck's teacher: Well, I didn't actually think he was daydreaming. I mean, he was so wrapped up in the squirrel and bird play going on outside the window that he didn't have time for daydreaming or for what was really important to me: math. And, yes, I was frustrated. I pulled out every trick in the book, and I couldn't seem to help the children make the connections they need to. And then zap! Chuck did it!

Critical colleague: What do you think was going on?

Chuck's teacher: Now that I can look back on it, I realize that he already knew the math I was trying to teach. I also know that my method of teaching wasn't working — but I knew that before!

Critical colleague: Tell me more about your method of teaching. What did you assume about teaching and learning with the methods you were using?

an interest in the educational outcomes of their practices. This reflects the democratic nature of action research, which avoids hierarchical distinctions among researchers, educators, and persons being educated.

Like qualitative research, action research is flexible. Research questions may be modified as data are collected and hypotheses emerge after the study begins. However, to qualify as research, action research must include a plan for systematically collecting and analyzing data. Decisions need to be made regarding:

- What types of data will be collected?
- How will the data be collected?
- Will measures need to be developed?
- What measures will best fit into the setting and not intrude on the usual practices?

Based on what they have learned in steps 2 and 3, action researchers then decide what they need to do to improve their practices. This might include modifications or refinements of lesson plans, new outreach to parents, development of

Chuck's teacher: That's an interesting question. I was caught between two worlds with my teaching. I hadn't thought about it consciously until you asked, but I was trying to teach the way I think children learn, with manipulatives and having the children work in small groups. But I was also trying to have a whole class discussion when I realized that most of the children weren't getting the concept as quickly as I thought they would be. So I went back to the way I was taught, the lecture.

Critical colleague: Were you caught between "encouraging their exploration" and "direct teaching"? Did you think you had to do one or the other?

Chuck's teacher: Precisely. In a sense, I didn't trust the newer way of teaching, although that is really where my intuition tells me we ought to be moving. I slid back into the way I was taught, and it didn't work. I was frustrated with myself, although I didn't know it at the time. I had a whole lot of "shoulds" buzzing in my head—"They should be getting it faster! They won't be ready for the proficiency tests!" *"Do something!"* And in the *"do something"* voice, I was the font of all knowledge; and if not all eyes were on me, well, how could they possibly be learning?

Critical colleague: And Chuck wasn't helping, was he?

Chuck's teacher: No. He was like a match to gasoline! And now I see that match was just what I needed! I was part of the problem; Chuck was part of the solution! (Arhar et al., 2001, pp. 99–100) new programs, use of new instructional strategies, or changes to existing school policies. This step distinguishes action research from other types of research and is an essential part of the process.

Most action researchers collect at least two different types of data and often include both quantitative and qualitative data. The measures should be appropriate to the setting and may include archival data that are already being collected, such as attendance records or report card grades. Data collection must also involve simple and short procedures that do not interfere with the normal activities in the setting. A classroom teacher who wants to conduct interviews might schedule these as part of parent-teacher conferences or conduct brief interviews with individual students while others are working in groups.

The research plan should include consideration of the length of time for data collection, determination of times when data will be collected, the frequency and places where collaboration will occur, and identification of methods to check on the quality of the data. Someone simply sitting down at a computer and reconstructing what happened in his or her classroom over the past month is not doing action research. Planning and systematic collection of data are essential to maintain the integrity of action research. Exhibit 12.2 presents questions suggested by Arhar et al. (2001) and Stringer (1999) that practitioners can use to plan an action research study and examples of how Yolanda might have answered them.

EXHIBIT 12.2
Questions for Planning an Action Research Study

- What is my research question or goal?

  Yolanda's research question was, "How can I improve my students' understanding of adding and subtracting negative numbers?"
- What can I try out in order to improve my practice?

  Yolanda tried a card game that she found during her search of the literature.
- How will I determine if there is a positive change in my practice?

  Yolanda gave students a quiz before and after the card game.
- How will I document the processes involved?

  Yolanda videotaped students playing the game and conducted student interviews.
- Do my data include sources taking both a broad and specific focus? Do my data reveal both the general picture of what is happening in the educational setting as well as information on the experiences of individuals or small groups?

  Yolanda's videotapes provided a picture of the whole class, but her interviews were with individual students.

Step 5: Carry Out the Action Plan

In step 5 action researchers carry out their action plans and continue to collect data and monitor the changes in student learning or school-based practices. This involves implementing the new strategies or intervention identified through reflection, collaboration, and the review of literature. The data collection methods should examine both the outcomes and processes involved in the implementation of the action plan. Note that Yolanda collected outcome data (test scores) and process data (observations of students and her own reflections).

Although action research usually involves activities that are part of normal educational practices, ethical issues may arise. These are discussed in Box 12.1.

Step 6: Analyze Your Data

Like qualitative researchers, action researchers analyze the data as they go. The types of data collected may be changed based on what they learn. In part, this reflects the fact that action research is embedded in an ongoing set of educational activities and is being conducted by a practitioner who modifies practices to fit the needs of learners. Action researchers do, however, review all of their data

- How does my plan promote reflection and include different viewpoints to ensure that the data and interpretations are credible?

  Yolanda reflected on her data with colleagues in her workshop. She also compared the videotape and student interviews to see whether her viewpoint matched that of the students. Finally, she examined the videotape and student interviews to see if they supported the evidence from the quizzes that student learning had improved.
- How will I analyze the data? Which types of data will be compared?

  Yolanda planned to calculate descriptive statistics on the quiz data. She also planned to code data from the observations and interviews and look for themes related to student learning. She planned to compare all three types of data to see if similar patterns emerged.
- How will I share my results with others to enhance educational knowledge and continue the cycle of action research?

  Yolanda shared her results with colleagues in the workshop and planned on continuing to use the game strategy with future classes to see if it was effective with other groups.

Source: Adapted from Arhar et al. (2001); Stringer (2008).

BOX 12.1
Ethical Issues in Action Research

Action research aims to improve education, so you might think there would be few ethical concerns involved. However, Pritchard (2002) points out that action research often involves sensitive ethical issues and questions. Action research has two entangled goals: improving practice and advancing educational knowledge. At times, these goals might result in conflicts. For example, a practitioner might wish to share with colleagues research findings considered important. To promote full understanding of the findings and research methods, extensive background and contextual details might be provided. However, this makes it difficult to maintain the confidentiality of the participants in a small educational setting. In addition, the consequences of revealing a participant's identity may be substantial, since the results of action research are often shared with the local community whose members are likely to know participants.

Consider the issue of research being voluntary and participants being free to decline participating. Practitioners such as teachers, counselors, administrators, and school psychologists often wield considerable power over students and parents. Although they may be told that their participation in a project is voluntary, they may feel unstated pressure to take part.

If the study involves interviews, videotaping, or unique interventions, action researchers must seek informed consent from parents and possibly students. However, since the researcher's major role is perceived to be helping students, participants may deemphasize the research goals and think that the informed consent is simply an invitation to receive a special and effective treatment. Although the informed consent might state that the project involves research to see *whether* an intervention is effective, participants might assume that they are being invited to participate *because* the intervention is known to be helpful. Practitioners, unlike applied researchers, have a high level of trust with persons likely to be participants, and they must be careful not to betray this trust.

at some point to see what it suggests about their practices. Both qualitative data analysis techniques (e.g., identification of themes and patterns) and descriptive statistics (e.g., graphs, means, and measures of variability) might be used to summarize the data.

One of the hardest aspects of action research is making time to complete the analysis of the research. Often data collection seems effortless because the data collected may involve procedures that a practitioner is doing anyway as a normal part of practice. However, reviewing the notes, numbers, journals, papers, and other types of data can be time consuming. It is often useful to review the

The flexible nature of action research also raises problems with institutional review boards (IRBs). Like qualitative researchers, action researchers often change methods during the data collection process. Although action researchers might discuss possible changes in their initial IRB proposals, it may be difficult to anticipate exactly how procedures might change during the study. It is therefore important to discuss with the IRB how much flexibility is allowed before the researcher is required to resubmit new procedures for review by the IRB.

Some of these problems can be alleviated if action research is participatory and democratic, with decision making and responsibilities shared by multiple educators, parents, students, and community members. If participants are actively involved in the project, misunderstanding of informed consent is less likely, modifications in research methods can be discussed, participation is more likely to be truly voluntary, and participants can have a voice in how and if information is shared. However, not all participatory action research projects produce consensus. There may be disagreements about how to use the results or how to modify an ongoing project. Critical action research often challenges the status quo. Therefore, it might be difficult to get approval for a study from major stakeholders (e.g., administrators or superintendents); without such approvals, an IRB might be reluctant to support a proposed study.

If you are planning an action research study, it is critical that you submit your project to your IRB as early as possible. You should think deeply about how you will address these ethical issues and provide as much detail as possible in your IRB proposal. Pritchard (2002) suggests that IRBs often need to be educated about action research since its methods differ from applied research. It may also be helpful to note how your study fits the framework of action research in your IRB proposal. As in all of action research, deep reflection about how you will balance your role as both a practitioner aiming to improve practice and a researcher aiming to advance knowledge is essential!

data with collaborators who can help to interpret what the data show about the problem and analyze any changes that resulted from the action plan. Descriptive statistics may be used to analyze numerical data and themes or patterns analyzed for qualitative data as discussed in chapters 3 and 5. Ultimately, however, the practitioner must decide what the meaning of the data is for practice. If students report that they are embarrassed when their parents come to class, should a teacher find roles for parent volunteers outside the classroom? Or should the teacher examine what parents have done in the classroom that could be changed to make it less embarrassing? Data analysis in action research actually occurs at all stages of the study, not only at the end of a cycle. You would not want to wait several weeks to examine student papers or parent surveys! Rather action researchers examine data as it is collected and reflect on what it shows. The action plan might be modified based on what the early data show. If a meeting with parents reveals that they are uncomfortable implementing a behavioral strategy at home, a new strategy might be selected and measures of how the plan is implemented would need to change. Flexibility is key to good action research.

Step 7: Form Tentative Conclusions

Although it is often hard to know when the cycle of action research is completed, at some point you will form tentative conclusions based on the analysis of your data. Did your action plan improve your practice? How did these changes occur? What questions remain? Action research is an ongoing process of reflecting on practice, identifying problems, and formulating new research questions and action plans to remedy the problems. So it never ends, at least as long as the practitioner feels there are still ways to improve practice. Once an action researcher has built a network of collaborators, it is easier to keep the process going.

DATA SOURCES FOR ACTION RESEARCH

All the research methods we've discussed have as their goal making data collection more systematic. However, because of the intense demands of action research, in which one individual takes on the roles of both researcher and practitioner, methods of data collection in action research must be simple and must fit into existing processes of the school or classroom. Most action researchers do not follow the complex procedures involved in formally establishing reliability and validity; however, we advise students to consider how they will demonstrate consistency and accuracy in their quantitative measures. Similarly, most action researchers do not conduct the type of lengthy observations or meticulous note taking typical of applied qualitative researchers. However, they do use reflective tools and critical collaborators to determine whether their qualitative data are credible and free from bias. Most action researchers collect data using several brief measures (rating scales, reflective journals, interviews). The data from these different measures are compared for consistency and accuracy.

Following we discuss some of the measures employed most frequently in action research studies. Some of these tools are discussed in more detail in chapters 4 and 5. We discuss here how they might be adapted for use in an action research study. Remember that data collection tools in action research can be used to examine and document three broad categories of data: our own behaviors, beliefs, feelings, and actions; the behaviors, beliefs, feelings, and actions of others; and physical settings, documents, or objects. Some of the tools described next can be used to analyze both ourselves and others.

Logs and Journals

Qualitative tools such as logs and journals are frequently used in action research because they are flexible and can be adapted to fit the setting. **Logs** with timed entries demand frequent recording, and these are usually completed by persons who can sit back and observe for extended periods (e.g., an aide). In action research, log entries might be brief, such as the sample log in Exhibit 12.3.

EXHIBIT 12.3
Classroom Log for Special Education Teacher

Log Entry for Week of January 19

Observer: Dean

Focal student: Aron

1/19/09

Hallway: Much activity and discussion in the hallway. Students wearing Obama buttons or stopping to examine the posters on our first black president. School has tried to use the election to promote pride in our diverse students. Aron enters looking at floor. He seems oblivious to the excitement.

Classroom: Aron is doodling as Ms. Fitts discusses the inaugural activities with the class. But when they talk about the dog that the new first family is getting, Aron looks up and says, "They should get a lab; they're the best!" He smiles and continues to draw. Later I see that he is drawing dogs.

Note to self: Aron is sometimes paying attention even when he seems not to be. Maybe pets could be an entry point to get Aron to talk with me more or to work on some skills during our take out time. I should find out if he has a pet and maybe try reading books on dogs with him.

1/20/09

Inclusion setting: I read a story on dogs with my three students with reading problems including Aron. He seemed to pay better attention than usual and asked questions that showed good comprehension of the story.

Note to self: I have to find other ways to use my students' prior knowledge and experiences in our activities. But then how do I maintain their interest and enthusiasm when the material is not personal?

Journals are another qualitative tool frequently used by action researchers. Teachers and other professional educators use journals to reflect on their feelings, perceptions, and interpretations of what they see happening in their classrooms or offices as part of their ongoing practice; these tools fit well into educational settings. An ongoing journal might be included as one of the data sources or, if no journal is in use, the researcher might introduce one for the study. In action research, journals may be completed individually or might be used to promote collaboration among participants. For example, Geer and Hamill (2007) used electronic journals to facilitate collaboration among preservice teachers in special education and general education who were completing field experiences for their teaching methods courses. Participants responded to weekly prompts such as the following:

- What is the climate of the classroom, and how do you know from your observations that this classroom contributes to individual student learning?
- What is the behavioral management plan in your classroom and why do you think it is or is not successful?
- How would you describe the way that you and your cooperating teacher are creating a diverse environment ensuring that all students have an opportunity to participate (that is, students with special needs, girls, minorities, different cultural groups)? (Geer & Hamill, 2007, p. 537)

The journal discussion provided a rich source of qualitative data on how both groups approached situations that arose during their field experiences and how they thought about and handled these. The discussion also promoted dialogue that enhanced the experiences of all students. Journals may be completed by any participant in a study, including teachers, staff, students, parents, or others. A sample of a structured student journal for an after-school science program is presented in Exhibit 12.4.

Field Notes

As discussed in chapter 5, qualitative observations are often recorded in the form of field notes. Action researchers rarely have time to record the extensive notes taken by qualitative researchers such as ethnographers; however, they may record brief notes to document what happens during a specific day, event, or activity. Sometimes data might be recorded only at predetermined intervals. For example, one might record the behavior of a focal child every 10 minutes. As in qualitative research, it is helpful to develop a simple protocol and recording sheet for these observations that provides space for recording both descriptive field notes and reflections. Exhibit 12.5 provides a sample of field notes organized to provide environmental context as well as descriptive and reflective field notes.

EXHIBIT 12.4
Structured Student Log

Student Journal Entry for After-School Mentoring Program

Date:

Where did you meet with your mentor?

What did you do?

Indicate if your activities involved any of the following by checking in the second column.

Generating ideas	
Forming hypotheses	
Collecting data	
Analyzing or thinking about data	
Reaching conclusions	
Reading	
Searching the Internet	
Working with other students and their mentors	

Pick one of the categories you checked above and explain what you did with your mentor that fit this category.

Other thoughts or comments?

Shadowing

In **shadowing,** the researcher accompanies a person during his or her normal activities for a full day. Throughout the shadowing, the researcher takes notes on what the person says or does but might also ask questions and interact with that person. Shadowing provides an in-depth slice-of-life look at a specific person or

EXHIBIT 12.5
Field Notes from Action Research Study

Observational Recording Sheet			
Student Observed: Kyle Grade 3 Age: 8 Gender: Male Date: February 22 Starting Time: 9:00 AM Ending Time: 11:45 AM Teacher: Ms. Lee (L) Student: Kyle (K) Graduate Student: Moira (M) works with K Observer: Kathy			
Time	**Student behavior in 10-minute intervals**	**Environmental context (What is happening in classroom? Room arrangement? Focus of class? Materials used?)**	**Reflections, impressions, hunches, interpretations**
9:15	M sits down with K who is looking around. M directs his attention to the problem, but he seems unclear what to do. She shows him how to break up the problem into parts. K says, "Oh, that's awesome." He picks up the pencil and completes the problem.	L writes Math for Today on the board and passes out worksheets with the problems. This is a daily routine and students begin working immediately. As students work, L checks their homework and touches base with students missing assignments.	

situation. Collaborators are critical when using shadowing as a tool because few practicing educators can afford to devote a full day to shadowing someone. It is a strategy that can be used by graduate or undergraduate students working with practitioners in the field. Or persons working together on an action research study might plan how to make time for shadowing. Data from a shadowing experience are recorded as field notes, using either a log format or an unstructured recording sheet.

9:25	K has completed his sheet and smiles at M. He raises his hand to answer L's questions. He does not get called on, but he still pays attention and checks off his problems as he gets them right.	L reviews math concepts from the previous week. One student was absent and she uses other students to help catch up. She asks them what "numeral" means and calls on one answer. She repeats the answer commenting that it is "fancy word for number."	The relationship with M seems an important part of K's motivation to do well. L said that M's work helps him to pay attention better because he understands the concepts before they work on the problems. That seems to be the case here.
9:35	K is somewhat distracted watching another student. L calls on him to get his attention and prompts him so he knows what she is asking. He doesn't know, and seems to pay better attention. When L calls on him again and he gets the answer right, he smiles broadly and raises his arms in celebration. M and L smile at him.	L invites students to share their answers to problems and has them write on the front board. Most raise their hands eagerly. She works through the problem step by step calling on different students to provide each part of the answer. She seems to try to have each student answer at least one question.	L uses firm but caring strategies to manage behavior in the class. She conveys her caring and most students, including K, seem motivated to do well. Over the months that M has worked with K he has shown a stronger interest in mastering math.

Checklists and Rating Scales

Action researchers frequently use checklists and rating scales as part of their teaching and their research. Teachers might design a checklist or rating scale for the strategies or skills that they hope to see in their students' writing and record information on the skills that they see, along with the dates or assignments on which the skills were shown. Exhibit 12.6 presents part of a checklist used at one school to record writing strategies of students.

Critical action researchers might view the use of researcher-controlled quantitative tools such as checklists or rating scales to be inconsistent with the democratic nature of action research. If they do use such tools, they would likely obtain reports from both the researcher and other participants, such as students

EXHIBIT 12.6
Checklist to Record Student Writing Strategies

Weekly Writing Checklist

Name:

Instructions: At the end of each class, put a tally mark next to any activity done in that session.

ACTIVITY	DATES							
	1/10	1/17	1/24	1/31	2/7	2/14	2/21	2/28
Prewriting								
Drafting/ Outlining								
Revising								
Editing								
Conferencing/ Talking								
Reading								
Other								

or parents. For example, data might be recorded on checklists by students who report their own skills and activities, as in Exhibit 12.6.

Checklists for action research might also be open-ended, with categories identified and space for recording information about student strengths and weaknesses. Students might be asked to report individually on their experiences, or the group might be asked to report on all of its members using the checklist. A sample of an open-ended checklist is presented in Exhibit 12.7. It is similar to an observational protocol but is more structured in that it identifies specific behaviors or questions to be addressed in entries.

Rating scales are an alternative way to record researcher or participant behaviors, observations, or attitudes. Simple rating scales might be designed to assess student reactions to assignments. For young children, these rating scales might even use nonverbal responses such as the one in Figure 12.2.

As noted in chapter 4, many standardized instruments utilize a checklist or rating scale format. Some action researchers, especially counselors, school psychologists, or administrators, often use these standardized instruments as part of their normal practice to assess strengths and weaknesses or record change over time. However, when tools are used multiple times to monitor behaviors, it may be necessary to simplify them so that they are suitable to the setting.

EXHIBIT 12.7
Checklist with Open-Ended Categories

Checklist for Group Problem Solving

Give examples of how your group completed each of the following steps in your problem-solving session.

Problem-solving step	**How we did it**
Defined our problem	
Collected information	
Brainstormed possible solutions	
Evaluated ideas and identified best possible solutions	
Tested possible solutions	
Finalized our decision	

FIGURE 12.2 Rating Scale Using Nonverbal Responses

How I felt about working in groups today
(Instructions could be read to child)

Interviews

Interviews are a type of qualitative data collection tool that can reveal participants' own perceptions of their views, feelings, or experiences. In action research, interviews are typically brief; they may occur during regular classroom or school activities in either a group or individual format. For example, in one study a kindergarten teacher briefly interviewed her students in class to find out how they felt about the hands-on literacy experiences that she was trying out (Brown, 1998).

While interviews are a good tool to use to find out how people feel about ongoing activities or changes in a school setting, they are vulnerable to inaccuracy and distortion. In action research, a teacher's or administrator's authority in the school may make participants feel uncomfortable revealing some thoughts or feelings. Involving others in planning a study and in data collection and analysis can promote trust and honest responses. One of our students who wanted to study cheating in his sixth-grade class worked collaboratively with his students, who suggested a reciprocal interview format in which he first interviewed them and then they interviewed him. He was amazed at the candid responses that the interviews produced and learned a great deal about how some of his assignments encouraged dishonest work.

Focus group interviews may fit naturally into a classroom setting in which students are already divided into teams or working groups. Some researchers have students lead and record data from group interviews. Arhar et al. (2001) described a study in which a teacher investigated student attitudes about environmentalism through focus group interviews led by students. The teacher and students first brainstormed questions and topics and then moved to one of five tables in the class for the interview. Discussions at the tables were tape recorded, and students at each table selected a person to facilitate the discussion and one person to record notes. One table had no tape recorder for students who did not want their comments to be recorded. The teacher visited each table briefly during the discussion but did not stay long. The analysis revealed some surprisingly negative views toward environmentalists, such as they were "nerds" or "did not bathe often!"

In designing interviews for action research, be sure to select approaches and questions that fit into your educational setting. Many of the guidelines for setting up interviews and writing interview questions discussed in chapter 5 are relevant here, but any techniques used must be acceptable to all involved in the study.

Surveys

When it is not possible to interview everyone, written surveys provide another way to find out what participants in a study think, feel, do, or value. In action research, surveys are often very brief and are more likely to include open-ended questions (despite our advice in chapter 8 not to overuse open-ended questions). The number of persons responding to a survey in action research is often small, so analysis of data from open-ended responses is not a problem. In addition, open-ended responses fit the democratic nature of action research by allowing persons to say what they think rather than selecting an answer from a list predetermined by the researcher. However, to reduce time required to respond, structured questions should be used when likely responses can be identified. Exhibit 12.8 provides an example of a simple teacher survey on use of technology that combines structured and unstructured questions.

While surveys used in action research are not likely to be subjected to the rigorous procedures described in chapter 8 for validating a survey, collaborators in action research often serve as a pilot group to critique a survey before it is distributed. Many of the types of questions and techniques for ensuring reliability discussed in chapter 8 may also be used in developing a survey for action research.

Maps

Maps depicting the layout of classrooms and schools are useful ways to communicate aspects of an action research setting that are hard to capture in words. A map can show at a glance how a classroom is laid out to facilitate discussion or how certain groups are isolated within a school. Maps may also assist a practitioner in thinking about the interactions, traffic flow, or organization of resources in a school or classroom and might help to visualize alternative arrangements. Anderson, Herr, and Nihlen (1994) described how a kindergarten teacher used a classroom map to record which resources children used at different times during the day. Finally, maps convey much about the context of a classroom, which is important information in this largely qualitative type of research.

Photographs, Videotapes, and Audiotapes

Photographs are another type of visual data that can capture rich detail about the people, activities, and context of an educational setting. Researchers might take photographs themselves or might ask participants to take photos using disposable cameras. This provides a useful window into the participants' experiences and

EXHIBIT 12.8
Teacher Survey on Use of Technology

This survey will help us assess what types of technology are most used and most needed at our school. Please answer questions as completely as you can and return the survey by September 30 to the Technology Survey box in the main office. If you have any questions, feel free to contact Len Ovo, Technology Coordinator in Room 108. Thank you for your time.

1. How often do you take your class to a computer lab? Check one.
   a. Every day
   b. 3–4 times a week
   c. 1–2 times a week
   d. 1–3 times a month
   e. 3–6 times a year
   f. 1–2 times a year
   g. Never
2. What activities or programs do students use on the computer? Check all that apply.
   - ❒ Searching the Internet
   - ❒ Writing papers
   - ❒ Creating slide shows
   - ❒ Creating graphics
   - ❒ Using spreadsheet
   - ❒ Using educational software programs: Please identify which programs are used:

     ____

perspectives and enables the researcher to share the products of the research—that is, the photos—with collaborating participants. Videotapes and audiotapes are more obtrusive and run the risk of changing participants' behaviors as students "mug for the camera." However, students do adjust to the presence of cameras and recorders that are left in place over time.

One advantage of videotapes and audiotapes is that they allow the researcher to record actions and interactions more fully and to revisit the events at a later time. This is a major advantage when the practitioner is trying to juggle the roles of researcher and practitioner. In addition, with the current emphasis on outcome-based performance assessment, videotapes and audiotapes allow the researcher-practitioner to document student knowledge, skills, interests, and motivation.

3. What types of technology are available in your classroom? Check all that apply and write in how often you use each.
   - ❒ Computers. How many? ____ How often used?

     ____

   - ❒ Projection system and screen. How often used?

     ____

   - ❒ Calculators. How often used?

     ____

   - ❒ Digital camera. How often used?

     ____

   - ❒ Smart board. How often used?

     ____

   - ❒ Other. Please describe:

     ____

4. What other types of technology would you like to have available in your classroom or the computer lab?

   ____

5. In what areas would you like to have training on the use of technology?

   ____

6. What other questions or comments do you have about our school's technology resources?

   ____

Thank you for your help. Please deposit your survey in the Technology Box in the main office! ☺

For example, Figure 12.3 depicts two youths and an instructor (K.H.V.) admiring a robot constructed by the youths in a technology-based summer camp. Notice how even in the still photograph, you can see the youths' interest and engagement and the complexity of their mutual creation. Photographs like this allow both the youths and the instructor to reflect on what they have learned and provide a means of communicating complex outcomes to other audiences.

Personal Records and Mementos

Personal records and **mementos** include letters, photographs, videotapes, and objects that form the personal record of an individual's life rather than those constructed for a research study. Arhar et al. (2001) suggest that in action

FIGURE 12.3 Photographs as Data Sources in Action Research

research, personal records and mementos can be used to promote deep reflection on one's own beliefs, values, identity, and experiences. A high school student might select photographs, letters, newspaper clippings, or even legal documents to tell the story of his family's immigration to the United States and how his values related to education have been formed. A teacher might look back on yearbook pictures of her own favorite teachers, college term papers, poems, or movies that fueled her desire to become a teacher. This type of very personal self-reflection can help an action researcher to identify some deeply held values and potential assumptions that might influence a study. Arhar et al. caution that this type of emotional reflection involves the ethical responsibility to protect the confidentiality of revelations by participants and think about resources that help some persons deal with the emotional aftermath from difficult memories.

Grading Rubrics

As educators, many of you may be familiar with grading rubrics. Rubrics attempt to provide standardization to the observational processes involved in grading student work. In the case of rubrics, the data that are being collected are often qualitative in nature (e.g., portfolios, papers, drawings, projects), but a quantitative judgment needs to be made regarding the data. This may be a numerical rating or letter grade. The rubric is essentially a set of goals or standardized criteria that are listed along with the various levels of attainment the student or portfolio might achieve: for example, high, medium, or low or achieved, developing, and beginning. Presented in Exhibit 12.9 is an example of a rubric that a teacher might use to assess a student's photography project.

EXHIBIT 12.9
Grading Rubric for Student Photography Project

Student:

Grade:

Title of Project:

Rate the level of skill shown in each using the following scales:

5 = very high

4 = high

3 = average

2 = low

1 = very low

Technical Skills

_____ **Quality of focus**

_____ **Skill in use of lighting and contrast**

_____ **Clarity of printing**

_____ **Quality of composition**

Creativity and Originality

_____ **Originality of concept**

_____ **Creative use of materials**

_____ **Overall creativity of presentation**

Records and Documents

Educational settings are inundated with paper and computer files! These records and documents can reveal much about the inner workings of a school. Minutes of meetings, report cards, attendance records, discipline records, teachers' lesson plans, letters from parents, and written evaluations of teachers are all part of the massive amounts of data available in records and documents. As noted in chapter 4, much of this archival data is also used by professional researchers. Action researchers are likely to utilize the records and documents produced within their immediate settings. These might include samples of student work, notes from meetings with parents or students, report card grades, syllabi, case records, calendars, assignments, or phone logs. The school or organization might also produce records or documents that can be used in action research, such as school yearbooks, newspapers, newsletters, memos, minutes from meetings, annual reports, databases of achievement data, budgets, or school Web sites.

Records and documents can be a valuable way to corroborate information from other sources. If a teacher reports in observational notes that student motivation increased as a result of a new instructional strategy, attendance records can be used to back up those claims. In addition, records from a classroom and school can be compared to see if achievement or attendance rates differ.

Action research may also generate new records or documents. Teachers, counselors, and school psychologists involved in an action research study might set up a study group to discuss their plans and analyze data. Notes from the study group would provide a record of the processes involved in the study.

Artifacts

Schools and classrooms are full of "stuff" that researchers call artifacts. **Artifacts** are objects used in the process of teaching and learning or products that result from the process of teaching and learning. Artifacts in educational settings might include desks, sample lesson plans, portfolios, textbooks, bulletin boards, a laser pointer, electronic smart board, manipulatives, athletic equipment, or anything else found or produced in schools. Artifacts are often examined and described in writing as part of one's observations. Some artifacts might be more systematically analyzed to see if they convey underlying assumptions about the nature of education. For example, textbooks might be examined to see how they portray gender roles, age groups, or ethnic groups. The types of notices posted on bulletin boards might be examined to reveal which groups and events receive the most promotion at a school.

Physical Traces

When we walk in soft grass, we leave footprints. Similarly, educational activities often leave **physical traces** to show that we have been someplace or used something. These physical traces might include wear on well-used library books, markings on textbooks, readings from a counter on how often a cabinet was opened or a Web page was accessed, fingerprints on frequently viewed display cases, or fraying of a toy that was well loved. An observant action researcher can make note of these physical traces as evidence that certain resources or parts of an educational setting were heavily utilized. Sometimes, physical traces might provide information that differs from official records. A librarian at a city library once pointed out to us that books on sexuality were not often checked out, but they were worn and tattered and frequently left out in the hidden study places of the library by interested but embarrassed students.

With the wide variety of data tools available, it can be overwhelming to decide which tools you need for a study. Exhibit 12.10 presents a sample planning sheet used by one student to design an action research study and consider what types of data he wanted to collect. You might consider designing a similar planning sheet for your own study.

Although action researchers often create tools adapted to their specific educational settings, it is often helpful to start with samples created by others. Many resources are now available through Web sites on action research that have collected samples of tools developed and used by practitioners. In addition, journals

EXHIBIT 12.10
Sample Planning Table for an Action Research Study

Type of Data	Collection Dates						Data Analysis
Quiz scores	9/18	9/25	10/2	10/9	10/16	10/23	10/25
Student writing samples	9/15	9/22		10/6	10/13		9/30 & 10/19
Homework	9/17	9/24	10/1	10/8	10/15	10/22	10/2 & 10/22
Small group conferences			9/29	10/5	10/12	10/19	10/25
Student journals		9/25				10/23	9/28 & 10/25
Teacher journal	daily notes	daily notes	daily notes	daily notes	daily notes	daily notes	9/28 & 10/25

devoted to action research often present useful tools used in published studies. We have listed resources that might be useful to students planning an action research study at the end of this chapter.

EVALUATION OF ACTION RESEARCH

Evaluation is important to action research, just as it is in all designs. The quality of the measures as well as the design of the study should be evaluated.

Evaluating the Quality of Measures Used in Action Research

As noted previously, action researchers typically adapt the measures used to fit the educational setting. Therefore, the measures are likely to be simpler than those used in applied research and less likely to require high levels of training or be supported by extensive prior research. A strength of action research is that multiple measures are used and examined repeatedly throughout the study to ensure that the measures provide the type of data needed to answer the research question. Since practitioners are professionals with valuable insider knowledge, they have the knowledge and skills to judge whether their measures are appropriate. The quality of the measures is supported by continual scrutiny by collaborators, critical reflection by researchers, and triangulation of data at multiple points throughout the study. In addition, action researchers may follow some of the same general principles outlined in chapters 4 and 5 to provide evidence that their measures are scientifically sound.

Although qualitative measures used in action research do not provide the same depth of evidence for credibility and dependability as those used in more traditional qualitative studies, action research employs protocols and recording sheets such as those appearing in the exhibits throughout the chapter. The democratic sharing of responsibility for developing measures, collecting data, and discussing results parallels the processes involved in member checks and participant reviews of data; by doing so it promotes credibility and dependability. Reflections in logs, journals, or field notes provide a mechanism for revealing and examining personal biases and assumptions. Collaborators frequently perform the same roles as peer debriefers and serve as an additional check on bias.

Quantitative measures used in action research are not typically subjected to the same level of scrutiny as standardized quantitative measures regarding reliability and validity; however, some of the procedures used in validating published measures can be used in action research. For example, one might examine the reliability and validity of the grading rubric presented in Exhibit 12.9.

Interrater reliability might be examined by having two different teachers rate all or a small group of portfolios using the rubric. The teachers might be collaborators on the study and might agree to rate portfolios for each others' classes. Consistency between the two sets of ratings could then be examined to see if high interrater reliability was obtained. Notice that the rubric includes an explanation of the criteria used to obtain the rating, which should promote high interrater reliability.

Validity for the rubric can be established to ensure that the rubric measures what it purports to measure. Sampling validity might be determined by comparing the assignments in the portfolios and the rubric to see whether the rubric covers all possible aspects of the assignments. If the portfolio is meant to show whether the student has mastered the state learning standards for a given time period, one might check to see whether the rubric covered all of the learning standards. Item validity might be assessed by asking another teacher, possibly a collaborator, to examine the rubric and determine whether each section of the rubric appropriately assesses the skills and knowledge required by the state learning standards.

Content validity of specific assignments and tests can be examined by other teachers or be related to a state's learning standards as evidence of validity. In addition, the repeated use of practitioner-made measures and their refinement over time to fit a particular school or classroom provide evidence to support the reliability and validity of measures used in action research.

Videotapes and audiotapes can be helpful in establishing reliability for action research measures. If an observational checklist is being used to gather data from a classroom activity, the activity can be videotaped. The recordings can at a later date be viewed and scored by a collaborator using the same checklist. The data from both the checklist used during the class activity and the checklist data from the videotape viewing can then be compared. Videotaping can be valuable in training members of a research team who will collect data. Such training is necessary to maintain consistency across scorers.

Evaluating the Design of Action Research Studies

From a pragmatic point of view, the evaluation of an action research study might be said to depend on whether it identified something that worked to improve education. However, as with other types of research, we must ask, "How do you know it worked?" The answer depends not only on the quality of the data collected (as discussed previously), but also on the strength of the study's design and procedures and whether analysis of the data provides strong evidence that change actually occurred.

In designing studies, action researchers often borrow methods from many different types of research. Because their goals include understanding and changing educational processes and outcomes, they include methods appropriate to survey research, experimental research, and case study research within a single study. Therefore, one can evaluate action research using some of the criteria discussed in chapters 6, 8, 9, and 10, always taking into account whether the criteria are appropriate for the setting. For example, if an action researcher developed a survey, it might be pilot-tested and response rates considered. However, the questions might be changed when the survey is used at a later date based on what the researchers learn in their first administration. If an action research study uses case study methods to investigate why a certain student has problems learning, multiple perspectives could be considered by including the student, his or her parents, and special education teachers or school psychologists working with the student as collaborative members of the research team. It is possible that some team members have previous knowledge of the focal student. However, rather than being a problem, this previous knowledge might help the researcher more quickly identify why the student is having trouble learning.

Considering whether action research meets some of the criteria for other types of research helps to build confidence in the scientific nature of its results. However, applying these criteria without considering the unique nature of action research would undermine its purpose. In addition, because action research often employs several different approaches in the cycles of data collection and analysis, one cannot simply apply criteria established for all of these other types of research to a single action research study. For example, Yolanda used experimental methods by including a pretest and posttest in her study, but she had no control group, since it would be unethical to deprive some of her students of a beneficial strategy. This could be viewed as a very weak study. However, if she continues to monitor her students' achievement through additional quizzes, she can provide long-term evidence to support the effectiveness of her active learning approaches. In addition, her collection of both qualitative and quantitative data provides more extensive support of how and if her strategy was effective.

In analyzing data, action researchers do not typically use the complex statistical procedures involved in applied research to analyze results obtained through quantitative data. As Yolanda did, they might use descriptive statistics to summarize and present their quantitative data. Similarly, practical action researchers do not engage in lengthy analysis of the conceptual and theoretical meanings behind their qualitative data to the extent that most qualitative researchers do (although critical action researchers do frequently engage in analysis of the larger social and cultural issues involved in their research). Action researchers often use procedures similar to qualitative researchers, such as coding and triangulation, to identify patterns and ensure that their results are dependable and credible.

Since action research usually involves collecting multiple types of data, triangulation is used to corroborate conclusions. Triangulation might involve comparisons of both quantitative and qualitative data. Arhar et al. (2001) suggest that triangulation should involve a three-way comparison of different types of data: data based on researcher perceptions, data based on participant perceptions, and data based on analysis of documents or work produced by participants. In Yolanda's study, she compared her perceptions of the videotapes to the interview data from her students and to the quiz data. All data provided evidence of a positive change.

Action researchers often seek to document not only *whether* change occurred but *how* it occurred. Therefore, data should be collected and analyzed to examine both the processes involved in the educational setting as well as the outcomes. Yolanda's videotapes might not have provided evidence for how her students' thinking changed. However, if she asked students to verbalize their thoughts as they played the game, more information about how the game helped them to understand the rules of adding and subtracting positive and negative numbers might have been obtained.

Several authors have argued that action research should be evaluated using its own criteria and not simply assessed using criteria established for applied research (Anderson et al., 1994; Creswell, 2003). They have proposed sets of criteria or questions to evaluate the design of action research studies. We have adapted these criteria and questions as:

- *Did the study identify and clearly define a problem or issue in practice that needed to be addressed?* This involves both identifying a clear research question and clarifying the issue or problem through extensive reflection and initial data collection.
- *Did the researchers challenge and test their own assumptions and interpretations throughout the study?* Evidence for this might include using multiple reflective measures or collaborative methods to examine and analyze date.
- *Did the researcher develop a logical plan and collect data in a systematic and dependable way?* Anderson et al. refer to this as process validity, which is supported through extensive planning, use of well-developed measures, and thoughtful design.
- *Were different types of data collected and triangulated or compared, and did the data represent different viewpoints and perspectives?* Multiple data sources and triangulation are important to ensure comprehensive documentation of processes and outcomes and to demonstrate what Anderson et al. call "democratic validity."
- *Did the researcher collaborate with others who had an interest in the problem?* To what extent did others engage in planning and decision making during the study? True

collaboration involves others in many aspects of the research and involves what Anderson et al. call "dialogic validity."

- *What evidence is provided that the research led to a positive change or a solution to a problem?* Did it improve the lives of others or empower them to make changes in their lives? Ultimately, action research should improve education (outcome validity) and, in the case of critical action research, should empower persons to continue to better their lives (catalytic validity).
- *Did the study contribute to the researcher's ability to reflect on his or her professional activities and to continue to improve his or her practice?* Is the researcher more aware of possible interventions or instructional strategies and how these might help particular students? Is the researcher likely to continue to reflect on how to improve practice?

Analyzing Data From Action Research

As we have emphasized throughout the chapter, data analysis in action research is ongoing and continuous. Action researchers need to stop at several points throughout the study and reflect on their research questions and the data that they have collected. Mills (2003) suggests that researchers ask themselves questions such as:

- Is my research question still answerable and worth pursuing?
- What do my data suggest about my practices? What is working? What is not?
- Are there data that I have left out? Are there other data that will help me better understand my practice?

Action researchers use many data analysis strategies that are similar to those discussed in chapters 7 and 11. Like qualitative researchers, action researchers collect multiple forms of data, and decisions must be made about how to organize and review the data.

Craig (2009) recommends setting up a **triangulation matrix** to identify which data will be used to help answer each subquestion. In this matrix, each subquestion is listed in the first column and the types of data being used to think about each subquestion are listed in the row for that question along with the type of data analysis to be performed. The matrix identifies which types of data should be compared during the data analysis. For an example of a triangulation matrix that Yolanda might have used, see Table 12.3.

Some of the data may consist of records normally kept as part of one's practice (e.g., grade reports, notes on assignments, results from assessments). However,

TABLE 12.3 A Triangulation Matrix

Subquestion	Data Sources to Be Analyzed and Compared		
1. How does student achievement change when a middle school math teacher switches from a rule method to a card game approach to teaching about positive and negative numbers?	Quiz grades	Observations by teacher	
2. How do students feel about the two approaches?	Interviews with students	Observations by teacher	
3. How does the learning process for students differ in the two approaches?	Interviews with students	Observations by teacher	Field notes from discussion with other teachers in her workshop

the researcher still must decide how and when to examine the data in relation to the subquestions. In addition, the data must be triangulated with other types of data. For example, Yolanda collected quiz scores, and she also examined these in relation to her observations of students playing the math game and to data collected from her interviews.

Qualitative data may be coded and analyzed for themes and patterns again using one's subquestions as a framework. The researcher looks for recurring phrases, concepts, or ideas and examines how these provide insight into the issues addressed in the subquestions. Due to the flexible nature of action research, during an action research study new patterns may emerge that were not anticipated in the subquestions. Arhar et al. (2001) have identified some questions for identifying patterns that may be useful in analyzing data from a variety of educational settings. These are presented in Table 12.4.

Once a pattern has been identified in one set of data, the researcher looks to see whether it occurs in other data sources. This review continues for each subquestion until all the data have been analyzed. Some action researchers use concept maps to display relationships among the themes that they have identified. An example of a concept map is displayed in Figure 12.4.

Quantitative data may be summarized using frequency tables or polygons. For example, the frequencies of responses to questions 1–3 in the technology survey presented in Exhibit 12.8 might be summarized as shown in Exhibit 12.11.

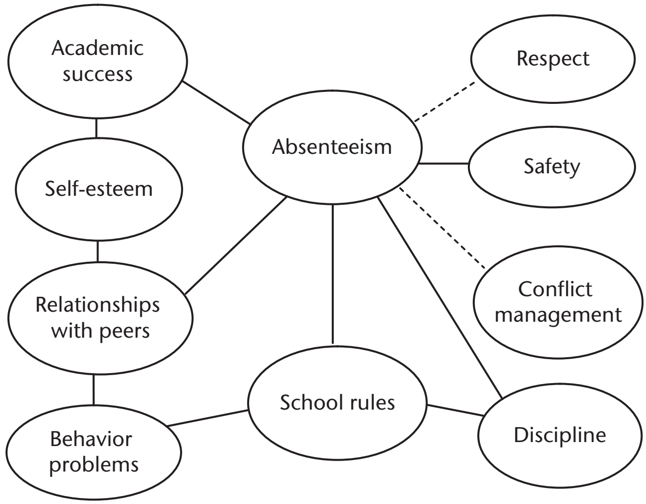
TABLE 12.4 Questions for Identifying Patterns in an Action Research Study

Type of Pattern	Questions to Ask During Data Analysis	Examples
Frequencies	How often does this occur? How often does "this" occur in comparison to "that"? Does it occur continually, at regular intervals, or only occasionally?	How often do children visit the book corner versus the art table during free-choice playtime? How often do French vocabulary words taught in class appear in student journals? How often do girls and boys seek help during science lab?
Times	What things occur simultaneously? Do similar things occur during similar periods or at similar times of the lesson, day, week, year? What sequences of behaviors or events occur?	Who arrives first in class each day and who leaves last? During what times are students most or least engaged? What is happening in class when student misbehavior occurs?
Spaces	How do people use space in this setting? What is close? Separated? Far apart? Above or below? Beside or opposite? Inside or outside? On or off? Over or under?	Where do the boys sit in class? The girls? The academically successful students? The academically unsuccessful students? Which teachers use the teacher lounge? Where are counselor or school psychologist offices located? What does the use of space in this setting say about what is valued?
Interactions	Who is interacting with whom? What are people reacting to? How do people interpret each others' actions?	Which students work together on the social studies project? How did students interpret my instructions on the journal entry? How do students from different cultures communicate with each other?

Causes and Effects	What are possible causes of this?	What are the possible antecedents and consequences of student misbehavior?
	What does this cause? What else changes because of this? Does this cause something else to increase? Decrease? What do I need to do to balance or unbalance this? Change or maintain this?	Why does a high school student have difficulty with algebraic equations but not have difficulty with probability?
		What changes occur in student belonging when students participate in school decision making?
		How does interdisciplinary teaming affect teacher planning and instruction?
Conditions	Is this a necessary condition for that to happen as it does?	When do White, Black, and Latino students socialize together and when do they socialize separately?
	Under what circumstances does this occur?	How much homework is needed to maintain steady progress in student writing skills?
Speed	Does this happen as fast as that?	Which methods of teaching keyboarding result in the fastest typing speed?

Source: Adapted from Arhar et al. (2001).

FIGURE 12.4 Use of Concept Map to Depict Relationships Among Themes



Source: Adapted from Mills (2003).

Descriptive statistics, such as mean, graphs, or percentages, might also be used to summarize quantitative data. For example, Sandall, Schramm, and Seibert (2003) used a histogram to display the results of a parent survey in their study of the use of literature to improve children's listening skills as shown in Figure 12.5 below.

Action research emphasizes collaboration, and data analysis often is conducted by groups of collaborators who meet together to review and analyze the data. When a group decides together on what the data mean, the consensus may facilitate implementation of action plans that involve multiple educators. Of course, data collection and analysis continue throughout and after the action plan is implemented.

After data have been analyzed, results are often shared with other educators. Sometimes results are shared only with others at one's school or program. However, increasing numbers of educators share their results in professional journals or in conference presentations. A rich body of literature on action research has accumulated; in 2010, there are more than 5,000 entries classified as action research in the Educational Research Information Center (ERIC). Journals devoted specifically to action research are listed in Box 12.2.

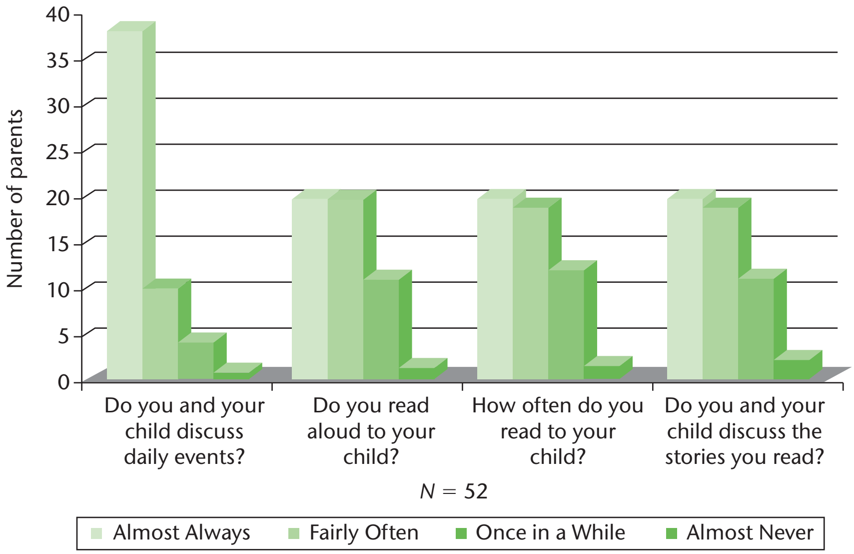
EXHIBIT 12.11
Summary of Frequency of Responses to Technology Survey

1. How often do you take your class to a computer lab? Check one.	
a. Every day	1
b. 3–4 times a week	1
c. 1–2 times a week	2
d. 1–3 times a month	5
e. 3–6 times a year	6
f. 1–2 times a year	2
g. Never	4
2. What activities or programs do students use on the computer? Check all that apply.	
☐ Searching the Internet	19
☐ Writing papers	11
☐ Creating slide shows	15
☐ Creating graphics	6
☐ Using spreadsheet	4
☐ Using educational software programs	12
Please identify which programs are used:	Mayaquest (2) Grammar Rocks (3) Bug Explorer (1) Oregon Trail (4) World Book Multimedia Encyclopedia (3) Spanish for the Real World (2) WriteOnline (3)

3. What types of technology are available in your classroom? Check all that apply and write in how often you use each.	How Many? Or Yes/No							How Often Used? (days per week)					
☐ Computers. How many? (top line) Number of responses (bottom line) How often used? (2nd column, top line) Number of responses (bottom line)	0	1	2	3	4	5	6	0	1	2	3	4	5
	2	6	6	3	2	1	1	3	4	5	2	3	4

☐ Projection system and screen. Yes/No How often used? (2nd column, top line) Number of responses (bottom line)	Yes = 8	1	2	3	4	5
	No = 13	2	2	3	1	1
☐ Calculators. Yes/No How often used? (2nd column, top line) Number of responses (bottom line)	Yes = 7	1	2	3	4	5
	No = 14	3	2	1	1	0
☐ Digital camera. Yes/No How often used? (2nd column, top line) Number of responses (bottom line)	Yes = 9	1	2	3	4	5
	No = 12	3	3	3	0	0
☐ Smart board. Yes/No How often used? (2nd column, top line) Number of responses (bottom line)	Yes = 5	1	2	3	4	5
	No = 16	0	1	1	1	2

FIGURE 12.5 Use of a Histogram to Display Results from a Parent Survey



BOX 12.2
Action Research Journals

- *AR Expeditions:* http://arexpeditions.montana.edu/index.php
- *Educational Action Research:* International print journal published by Taylor & Francis.
- *Networks: An On-Line Journal for Teacher Research:* http://journals.library.wisc.edu/index.php/networks
- *The Ontario Action Researcher:* http://www.nipissingu.ca/oar/
- *Wisdom of Practice: An On-Line Journal of Action Research:* http://www.vancouver.wsu.edu/staff/ludwig/ActionResearch/articles.htm

Finally, several organizations and universities have action research centers that provide resources to support action research and present samples of their student projects. These are listed in Box 12.3. We hope that many of the students reading this book will become more actively involved in this rich action research movement!

BOX 12.3
Action Research Centers or Institutes

Many of these research centers are located at universities that require students to complete action research projects as part of their studies. They often provide sample studies or data collection tools on their Web sites.

- Aston University, UK: http://www.cvar.org.uk/
- Brown University: http://www.lab.brown.edu/
- Center for Collaborative Action Research at Pepperdine University: http://cadres.pepperdine.edu/ccar/index.html
- Illinois Wesleyan University: http://www2.iwu.edu/action/
- New York University: http://www.wagner.nyu.edu/leadership/
- Queens University: http://educ.queensu.ca/~ar/index.html
- Research for Action Organization—affiliated with University of Pennsylvania: http://www.researchforaction.org/index.html
- Southern Cross University: http://www.scu.edu.au/schools/gcm/ar/arhome.html
- University of California at Los Angeles: http://www.college.ucla.edu/up/ccl/aboutus.htm
- University of East Anglia, UK: http://www.uea.ac.uk/care/ http://www1.uea.ac.uk/cm/home/schools/ssf/dev/odg/research/current-projects/SARC

SUMMARY

Action research is a type of research typically conducted by educators in their own work settings for the purpose of enacting immediate changes in an educational setting. Action research borrows techniques and ideas from all other types of research but is more fluid and flexible. It has a long history in many fields other than education. There are two types of action research, critical action research and practical action research. While action research is quite varied, it usually involves educators reflecting on their own practice, systematic planning and several cycles of data collection, taking action to improve practice, and collaboration with others.

The steps for conducting action research are fluid and flexible. Many different types of data can be collected in action research, including both quantitative and qualitative data. Measures may include data from observations, interviews, documents, records, visual images, surveys, personal mementos, classroom assignments, or artifacts. A major consideration is that the measure fit into the educational setting naturally. Action research should be evaluated based on criteria appropriate for the setting. The quality of the measures in action research is supported by scrutiny by collaborators, critical reflection by researchers, and triangulation of data at multiple points throughout the study. In addition, action researchers may follow some of the same general principles outlined in chapters 4 and 5 to provide evidence that their measures are scientifically sound.

While some authors evaluate the design of methods used in action research by applying relevant criteria from quantitative and qualitative methods, other authors argue that action research should be evaluated based on its own criteria. Criteria specific to action research emphasize clear identification of a problem, control of researcher bias, collaboration with others, systematic planning and data collection, documentation of processes and outcomes in the study, and triangulation of multiple sources of data. Data from action research studies are analyzed using strategies similar to those used in both qualitative and quantitative research studies. Some approaches recommended for action research data analysis include creation of a triangulation matrix, use of concept maps to analyze themes, and meetings with collaborators to analyze data. Results from action research studies are presented in either general professional journals or journals devoted to action research. Many resources on action research are available through professional organizations and universities with centers or institutes on action research.

KEY CONCEPTS

action research
artifacts
critical action research or critical pedagogy
journals
log
physical traces
personal records and mementos
practical action research
shadowing
triangulation matrix

DISCUSSION QUESTIONS OR ACTIVITIES

1. Using the questions proposed by Arhar et al. (2001) in step 2 of the action research process, identify a problem or issue in your own practice that could be examined using action research. Reflect on how your values and experiences have contributed to your thinking about this problem. If possible, discuss the questions with a peer who can serve as a critical friend.

2. Plan an action research study examining the literacy strategies used by middle school or high school teachers. Discuss several types of data you might collect, what your role in the research would be, how you might manage your role as a practitioner and researcher, and who might be your collaborators. What do you think are the advantages and barriers to this type of research?

SAMPLE ACTION RESEARCH STUDIES

Hashey, J. M., & Connors, D. J. (2003). Learning from our journey: Reciprocal teaching action research. *The Reading Teacher, 57*(3), 224–232.

Jordan, L., & Hendricks, C. (2002). Increasing sixth grade students' engagement in literacy learning. *Networks, 5*(1). Article 1. Retrieved June 26, 2009, from http://journals.library.wisc.edu/index.php/networks/issue/view/19

Keeling, K., Smith Myles, B., Gagnon, E., & Simpson, R. L. (2003). Using the power card strategy to teach sportsmanship skills to a child with autism. *Focus on Autism and Other Developmental Disabilities, 18*(2), 103–109.

SUGGESTED READINGS

Anderson, G. L., Herr, K., & Nihlen, A. S. (1994). *Studying your own school: An educator's guide to qualitative practitioner research*. Thousand Oaks, CA: Corwin.

Arhar, J. M., Holly, M. L., & Kasten, W. C. (2001). *Action research for teachers: Traveling the yellow brick road*. Upper Saddle River, NJ: Prentice Hall.

Short, D. J., & Echevaria, J. (1999, December). *The sheltered instruction observation protocol: A tool for teacher-researcher collaboration and professional development*. (EDO-FL-ii-09). Reston, VA: ERIC Clearinghouse on Languages and Linguistics.

Stringer, E. (2008). *Action research in education (3rd ed.)*. Upper Saddle River, NJ: Pearson.

Wadsworth, Y. (1998, November). What is participatory action research? *Action Research Inter-national*. Paper 2. Retrieved Sept. 10, 2004, from http://www.scu.edu.au/schools/gcm/ar/ari/p-ywadsworth98.html

CHAPTER THIRTEEN

PROGRAM EVALUATION IN EDUCATION

CHAPTER OBJECTIVES

- Define program evaluation and note several characteristics that separate program evaluation from educational research
- List several differences between formative and summative evaluation and describe the purpose of each
- Define the roles of an internal and an external evaluator, and describe some of the benefits and challenges that an evaluator working in these two different methods would encounter
- Describe the different approaches to program evaluation and discuss the similarities and differences in the approaches
- Describe the basic steps in conducting a program evaluation

RESEARCH VIGNETTE

An urban school district receives a 3-year grant to implement an after-school program to improve student academic achievement. As the district starts to implement the program, the district administrator realizes that an evaluation of the program is required. The administrator also realizes that such work requires the expertise of someone from outside the district, and the superintendent, with permission from the school board, hires an external evaluator from a local college. After reviewing the grant, the evaluator conducts an initial review of the program's curriculum and activities. Next, the evaluator develops an evaluation plan and presents it at the next school board meeting. The evaluation plan summarizes the objectives that the evaluator has developed and the tools that he will use to collect the data. As part of the data collection process, the evaluator discusses how the plan will provide two different types of feedback: formative and summative evaluation. Formative evaluation will be used to address issues as the program is happening. A question might be, *Are all the stakeholders aware of the program and its offerings?* Summative evaluation will be used to answer the overall evaluation question: *Did students in the after-school program have a significant increase in their academic achievement over those students who did not participate?* The board approves the plan, and the evaluator spends the next month collecting data for the formative and summative portions of the project.

At the next board meeting, the evaluator presents some of the formative evaluation and reports that there is a need to increase communication with parents. He suggests that the program increase the number of fliers that are sent home, update the school Web site, and work more collaboratively with the parent council. In addition, he notes that there is a wide variation in parent education levels within the district and that a large number of parents speak Spanish as their native language. The evaluator recommends that phone calls be made home to parents and that all materials be translated into Spanish. The district made changes to the project based on the evaluator's recommendations.

At the end of project year 1, summative findings are presented in a final report. The report showed that lack of parent communication was still a problem and there was little difference in the scores on the standardized measures used to gauge academic achievement for those students who participated in the program versus comparable students who did not participate.

Based on the evaluation report, district officials decide to make modifications to the program for the upcoming year. A parent center, which was not part of the original plan, was added to help increase parent involvement. In addition, the administration decided to cut back on the number of extracurricular activities the after-school program was offering and to focus more on tutoring and academic interventions, hoping that this would increase academic achievement in year 2.

WHAT IS PROGRAM EVALUATION?

Program evaluation examines programs to determine their worth and to make recommendations for refinement and success. A *program* is a set of specific activities designed for an intended purpose with quantifiable goals and objectives. While a research study could certainly examine a program, most research tends to focus on either generalizing findings back to a wider audience (e.g., quantitative research) or discussing how the study's findings relate back to the literature (e.g., qualitative research). Most research studies, especially those that are quantitative, do not focus on knowing how just one after-school program functioned in one school building or district. However, for those conducting program evaluations, this is seen as precisely the purpose.

Programs come in many different shapes and sizes, and therefore, so do the evaluations that are conducted. Educational programs can take place any time during the school day or after. For example, programs can include a morning breakfast/nutrition program, a high school science program, an after-school program, or even a weekend program. Educational programs do not have to occur on school grounds. An evaluator may conduct an evaluation of a community group's educational program, or a program at the local YMCA or Boys & Girls Club.

How Does Program Evaluation Differ From Research?

While program evaluation uses the same methods found in applied research studies, there are six ways that procedures used by evaluators are different.

First, program evaluation is used for decision-making purposes, while research builds our general understanding and knowledge on a particular topic and to inform practice. The different uses of program evaluation and research affect how the studies are critiqued and judged. An applied research study is often judged by how the data were collected and analyzed. Generally, the study is not critiqued on whether it made direct and immediate impact on what was being studied. The same is not true in program evaluation. A program evaluation report is judged both on its data collection methods and whether the information was used by those who are in the position to make programmatic changes. An evaluation report that is methodologically sound but "shelved" and creates little or no change is considered not to be of value. The program evaluator has a complex role that includes collecting data about programs as well as simultaneously working and building a relationship with others so that the final report will ultimately be used to improve the program. Program evaluation is similar to action research in that both aim to create immediate change. However, an evaluator is not usually involved in the day-to-day delivery of services or activities that are being evaluated.

A second difference is how the program evaluator and the researcher gain access to the project and program site. As described in the vignette, the program evaluator was hired by the school district to conduct the evaluation of their after-school program. In general, a program evaluator enters into a contractual agreement either directly or indirectly with the group whose program is being evaluated. This individual or group is often referred to as the **client**. As a result of this relationship between the program evaluator and the client, the scope of the evaluator's assignment can be restricted by the client. A client dictating what one would investigate for a research study would be very unusual. For example, because of the nature of qualitative research, a qualitative researcher who enters a school system to do a study on school safety may find a gang present in the school and choose to follow students as they try to leave the gang. If a program evaluation was conducted in the same school, the evaluator may be aware of the gang and students trying to get out of the gang. The evaluator might view this as an interesting phenomenon, but he or she would not pursue it unless the client perceived it as an important aspect of school safety or unless gang control fits into the original objectives of the program.

A third difference was demonstrated in the vignette: program evaluation often collects two different forms of evaluation data, **formative** and **summative**. The purpose of formative data is to "change" or "make better" the thing that is being studied (at the very moment when it is being studied). This is something that is typically not found in most applied research approaches. Rarely would the researcher have this "reporting relationship" in which formative findings were reported back to stakeholders or participants for the purposes of immediately changing the program.

A fourth difference is the speed at which program evaluation can change practice. For example, in a previous chapter the overall purpose of applied research (e.g., correlational, case study, experimental) was discussed as expanding our general understanding or knowledge about the topic and ultimately to inform practice. While this is certainly a main purpose of applied research, empirical evidence supporting a new method or approach doesn't mean that people will suddenly abandon what they have been doing for years and switch to the research-supported approach.

Notice how in the vignette change occurred very rapidly through the use of program evaluation. Based on the evaluation report, administrators, school board members, and project staff decided to reconfigure the structure of the after-school program and to provide parents with a center in the hope of increasing parent involvement. In addition, they decided that many of the extracurricular activities would be dropped and that the new focus would be on the tutorial component of the program. Those who implemented the recommendations made in the evaluation hoped to see even more improvement in students' academic scores in the coming year.

In contrast, take, for example, applied research in the area of instructional methods in literacy. In the 1980s, the favored instructional approach was whole language; however, a decade of research began to support an older approach called *phonics*. Despite the mounting evidence in favor of phonics, it took approximately a decade for practitioners to change their instruction. In the early 1990s, however, researchers began to examine the benefits of using both whole language and phonics in what is referred to as a *blended approach*. Again, despite substantial empirical evidence, it took another 10 years for many practitioners to use both approaches in their classrooms. Although this is a simplified version of what occurred, the purpose here is to show the relationship between applied research and practice and the speed (or lack of speed) with which settings or systems change as a result of applied research.

While there are certainly many program evaluations in which change does not occur as swiftly (or at all), one difference between program evaluation and research is the increased emphasis program evaluation places on such change occurring. In fact, there are philosophies and approaches in program evaluation that emphasize the utilization of evaluation findings so much that evaluators believe that if the evaluation report and recommendations are not used by program staff to make decisions and changes to the program, then the entire evaluation was a complete waste of time, energy, and resources (Patton, 1997).

A final difference between research and program evaluation is the way that program evaluation findings are presented. In conducting empirical research, the researcher commonly writes a study for publication—preferably in a high-level refereed journal. In program evaluation, and as in the vignette, the findings are presented in what is commonly referred to as the *evaluation report*, and not as a publication in a journal. (This is why students rarely encounter program evaluations when reviewing literature.) In addition, the majority of evaluation reports are given directly to the group or client that has hired the evaluator to perform the work and are not made available to others.

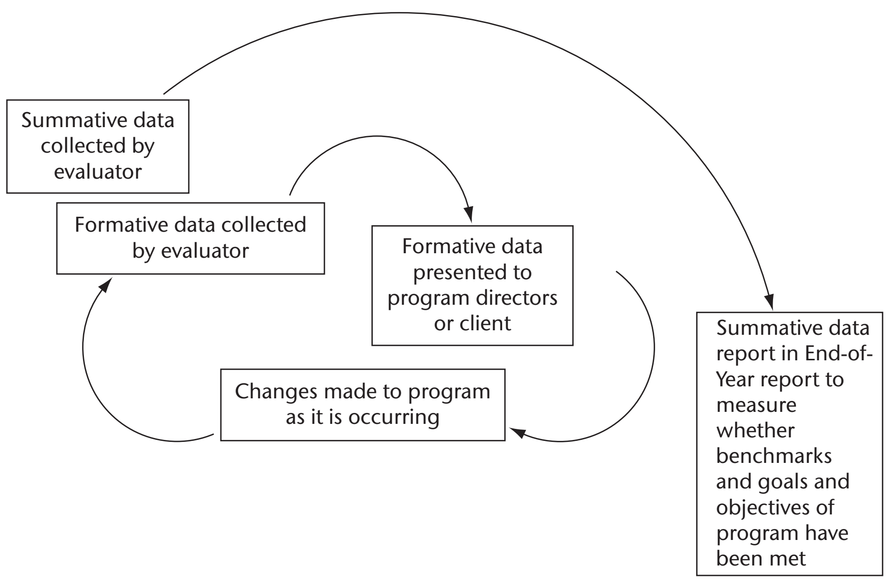
Formative and Summative Evaluation

Both quantitative and qualitative data can be collected in program evaluation. Depending on the purpose and the audience of the evaluation, an evaluator may conduct an evaluation that is quantitative or qualitative or a mixed-methods approach. In addition to using quantitative and qualitative data, a program evaluator has the option of providing summative and formative evaluation within a project (see Figure 13.1).

Summative and formative evaluations are not exclusively dictated by whether the evaluator collects quantitative or qualitative data. Many people have the misperception that summative evaluation uses exclusively quantitative data, and formative evaluation uses qualitative data. This is not always the case. Whether evaluation feedback is formative or summative depends on the type of information and when it is provided to the client.

Data for summative evaluation is collected for the purposes of measuring outcomes and determining how those outcomes relate to the overall judgment of the program and its success. As demonstrated in the vignette, summative findings are

FIGURE 13.1 The Summative and Formative Process in Program Evaluation



provided to the client at the end of the project or at the end of the project year or cycle. Typically, summative data includes student scores on standardized measures such as state assessments, intelligence tests, and content area tests. Surveys and qualitative data gathered through interviews with stakeholders may serve as summative data if the questions or items are designed to elicit participant responses that summarize their perceptions of outcomes or experiences.

For example, an interview question that asks participants to discuss any academic or behavioral changes that they have seen in students as a result of participating in the after-school program gathers summative information. This information would be reported in an end-of-project year report. However, an interview question that asks stakeholders to discuss any improvements *that could be made* to the program to better assist students in reaching those intended outcomes would be formative.

Formative data is different from summative in that instead of being collected from participants at the end of the project to measure outcomes—as in summative data—formative data is collected and reported back to project staff as the program is taking place. Data gathered for formative evaluation is reported back to the client in a timely manner. There is little value in formative evaluation when the evaluator does not report such findings to the client until the project is over.

Formative evaluation can be reported through the use of memos, presentations, or even through phone calls. The role of formative feedback is to identify and address the issues or serious problems in the project. Imagine if the evaluator in our vignette did not report back formative findings regarding parent communication. Imagine how many students might not have been able to participate in the after-school activities. One of the roles of the evaluator is to identify these program barriers and to inform program staff so that changes can occur. When programs are being implemented for the first time, formative feedback is especially important to developers and staff. Some programs require several years of intense formative feedback to "get the kinks out" before the program can become highly successful.

Formative feedback and the use of that information to change or improve the program separate program evaluation from most types of applied research. Remember the experimental or quasi-experimental research approaches discussed in chapter 9 and how the researcher tries to control extraneous variables so that only the independent variable can affect the dependent variable. An important aspect of experimental research is a clear definition of the different treatments or level of independent variable. If the program is the treatment variable, then it must be designed before the study begins. An experimental researcher would consider it disastrous if formative feedback were given because the treatment was changed in the middle of the study. By contrast, program evaluators, while trying to keep the independent variables or treatment constant, realize that it is better to make modifications to the program even if it "distorts" the lines of causality than to deliver a substandard program for the entire school year.

Internal and External Evaluators

The proximity of an evaluator to the subject of evaluation plays a critical role in access to information, collection of that information, and the reporting and utilization of that information to promote change. Just like the waiter at a restaurant (internal evaluator) has a different perspective of the food and the management from that of the food critic (external evaluator) who comes to dine and to write a review for the local paper, an evaluator's perspective and his or her relationship to the setting or program needs to be considered. While the food critic raves about the quality of the food, the waiter knows that the restaurant takes shortcuts in preparation and uses inferior grades of meat. As is obvious from this example, an **external evaluator** is someone from outside the immediate setting who is hired to "come in" and evaluate the program. Since this person has no obligations, he or she would in theory have no immediate biases for or against the program or any one of the stakeholder groups involved in the project. Most programs that receive federal, state, or foundation funding require an external evaluator to be present.

On the other hand, many companies, agencies, institutions of higher education, school districts and other groups employ internal evaluators. An **internal evaluator** is typically someone who is an employee of the company, agency, or group who is responsible for carrying out duties that pertain to evaluation. For example, many school districts now have a program evaluator on staff. The responsibilities of this person are to establish and work with databases to maintain student academic and behavioral data and to use data to assist staff and administration in improving practice. Internal evaluators at districts provide expertise in working with the state testing and accountability data as well as monitoring programs that the school is currently implementing.

There are many strengths and weaknesses to both internal and external evaluators. As mentioned previously, the main reason that many funding agencies require an external evaluator to be present is to increase the objectivity of the data that is being collected. However, the role of the external evaluator also has some weaknesses. External evaluators are often faced with the difficulty of establishing trust with the stakeholders involved in the program that they are evaluating. Even though the external evaluator is collecting data on the program and not specifically on the performance of program staff, this stakeholder group may not welcome the evaluator with open arms. Stakeholders may, and often do, see the evaluator as a threat—someone whose job it is to find "holes" in the program. They may see the evaluator's work as a direct threat to their livelihoods. The stakeholders may feel that the external evaluator "really doesn't know them . . . " or "doesn't know what they are all about. . . . " They may feel that the evaluator doesn't know enough about the setting or the context of how things work in that setting to be able to gather in-depth data that would pertain to them and be meaningful for evaluation purposes. Stakeholders who are unsure and uncertain about this evaluator are likely to avoid the evaluator altogether by not returning phone calls to set up interviews or returning surveys. Entering a new setting, establishing trust with the various groups involved in the program, and providing meaningful data back to participants for programmatic improvements is a daunting and often difficult challenge for even the most seasoned of program evaluators.

On the contrary, internal evaluators typically do not have to deal with establishing trust from stakeholders as external evaluators do. In addition, internal evaluators know the setting, how to access needed data, and the "language" that each group uses. However, the internal evaluator's previous and ongoing relationships with persons in the setting may interfere with the evaluation. For example, an internal evaluator may be intimidated by a supervisor into distorting the findings from a recent project.

In some cases there might be both an internal and external evaluator. If an internal evaluator is already present in a school system or program, then an evaluation plan should encompass the work of both evaluators to optimize the breadth and depth of data that is collected and to ensure the overall success of the program. In such situations, the internal evaluator is responsible for collecting certain types of data that the external evaluator does not have to access. In turn, the external evaluator would collect additional data to ensure the authenticity and objectivity of the evaluation effort and its findings. For a more in-depth description of how an external evaluator works with clients, see Box 13.1, which talks about an evaluator's experience with expanding a science program to another state.

BOX 13.1

Train Them and They Will Come. . . . Well, Maybe: The Role of the Program Evaluator to Deliver Bad News

A program evaluator was hired by a group of curriculum developers to evaluate the expansion efforts of an inquiry-based science program for high school students. For the past 3 years the developers had been piloting their program on the east coast through a grant they received from a national sponsor. The following summer, the developers held a 3-week training for 30 teachers in the expansion state on the west coast. The program evaluator arrived the last few days of training to observe, administer a survey, and hold several focus groups with the teachers. The purpose of the site visit was threefold: to document the training, to gather teachers' perceptions about the quality of the training, and to document whether the teachers felt prepared to implement the program in their school that coming fall.

Following a break in the training, the evaluator administered the survey to the teachers-in-training. During lunch, he quickly scanned the surveys, and much to his surprise discovered that all the teachers indicated that they would *not* implement the program that coming fall. This was a surprise to the evaluator, especially since in the pilot all teachers had indicated that they would implement the program following the training. Believing that surely the teachers had misread the question, the evaluator adds this item to his interview protocol. During the three focus groups, all the teachers validate the survey finding and reaffirm that they are not going to implement the program that school year. Confused, the evaluator asked the teachers why they were not going to implement and found out that under their state education department, all new curriculum adopted by school districts has to undergo a yearlong review. This was not the case on the east coast, where school districts could adopt any curriculum desired by the administration.

The evaluator then asked the teachers if they had relayed any of this information to the developers during the previous 3 weeks. The teachers said that they had not told the developers because they thought they were nice people and didn't want to upset them. At the end of the day, the evaluator broke the bad news to the developers. The developers were shocked; they couldn't believe that the teachers weren't going to implement the program and had sat through the training without telling the developers. Having learned this, the developers changed the last day of the training to focus on what teachers could do to get ready for the program over the course of the next year. The teachers found the switch by the developers to be very helpful, eased their bad feelings about not being able to do the program in the fall, and all the teachers implemented the program the following year with great success. The developers and evaluator learned an important lesson: not to assume that what works well in one setting will automatically work the same way in another, despite how similar the two settings may appear to be on the outside.

TYPES, APPROACHES, AND MODELS OF PROGRAM EVALUATION

Just as researchers can choose among many types of applied research approaches, program evaluators can utilize different evaluation approaches. The model used depends upon the type of project, the needs of the stakeholders, and the desires of the client.

Objective-Based Approach

The most common approach to program evaluation is the **objective-based approach,** which is guided by objectives written by the creators of the program and the evaluator. Evaluation objectives are written statements that describe the overarching purpose of the evaluation and clearly state the type of information that will be collected. Many times, these objectives are supported further through the use of **benchmarks**. A *benchmark* is more detailed than an objective in that it specifically states the quantitative goals that the participants in the program need to reach in order for the program to be successful. (Note that these terms are used differently in evaluation than they are used in curriculum.) Presented in Box 13.2 is an evaluation objective followed by a program benchmark.

Many times evaluators start with the objectives for the evaluation and build evaluation data collection activities from those objectives. Evaluation objectives may guide either formative or summative data collection. Either way, quantitative and/or qualitative data is collected and findings compared to the project's objectives. While objectives are helpful in shaping the evaluation, evaluators may become so focused on the objectives that they lose sight of unanticipated outcomes or benefits to participants as a result of the program.

BOX 13.2
Examples of Evaluation Objective and Benchmark

Evaluation Objective: To document middle school student changes in academic achievement, particularly in the area of reading and literacy skills.
Benchmark: Students in grades 5–8 will show a 10% gain on the ELA—English Language Arts—state assessment in year 1 and a 20% increase in students passing the ELA in program years 2–3.

Goal-Free Evaluation Approach

The **goal-free evaluation** approach doesn't use evaluation objectives. This approach is guided by the perspective that there are many findings and outcomes that do not fall within the "tightly packaged" goals and objectives established by the project directors and the evaluator. Those who practice goal-free evaluation believe that the "unforeseen outcomes" are perhaps more important than the outcomes that the program developers envision. One of the difficulties in conducting a goal-free evaluation is that projects that receive funding are required to show specific outcomes based upon objectives. If the outcomes are not included in the evaluation, the appropriate data may not be collected.

Expertise-Oriented Evaluation Approach

The **expertise-oriented evaluation** approach is one of the oldest and most utilized methods of program evaluation. The evaluator is a content expert as well as serves a role as judge in addition to evaluator (Fitzpatrick, Sanders, & Worthen, 2004). Agencies granting accreditation to institutions, programs, or services send program evaluators to these sites to conduct expertise-oriented evaluations. In these situations, data are typically not collected by the evaluators but are "presented" by those participants being judged or seeking accreditation. The evaluators in this approach judge the program or service based on an established set of criteria as well as their own expertise in the area. An example of this type of evaluation is the National Council for Accreditation of Teacher Education (NCATE). Colleges and universities who train teachers often seek national accreditation in order to validate the quality of their programs.

Participatory-Oriented Evaluation Approaches

Participatory-oriented evaluation approaches take a very different perspective on program evaluation than the approaches described so far in this chapter. Where the focus of the previous approaches has been on the program and examining different aspects of the program, the participatory-oriented evaluation approach is ultimately interested in those who the program serves. Under this model, an evaluator would involve program participants in all aspects of the program evaluation. In some cases the evaluator trains the participants, who then set objectives, develop instruments, collect data, analyze data, and report findings.

Logic Models

Program evaluators are faced with many challenges when evaluating the worth or effectiveness of a program. Applied researchers oftentimes address these challenges through a research design that allows them to make such comparisons (e.g., a pre-post control group design). While these designs are beneficial for showing causality, they present logistical problems for program evaluators. Programs are typically developed to serve individuals and groups, not to exclude them; therefore, providing an after-school enrichment program to half of the student body while denying participation to the other half is not a common practice.

In the past 10 years program evaluators have turned toward logic models to help provide a "causal connection" between a program's activities and participants' meeting outcomes. **Logic models** provide a framework or template for clearly identifying what needs to be measured at each distinct stage of the program (Frechtling, 2007).They function on the assumption that a logical chain of events must occur in order to reach the final results of the program. An example of a basic logic model is presented in Figure 13.2.

Presented in Table 13.1 is a more in-depth description of the components that make up a logic model.

Program evaluators use logic models in many different ways. Logic models can be used in the initial stages of constructing an evaluation of a program. They are useful tools to "spark" a dialogue between developers of programs and those individuals the programs are serving. They can be applied to a program that is already being implemented to make sure the program is on track and that the activities being delivered will result in the intended end outcomes for participants.

FIGURE 13.2 Overview of a Logic Model

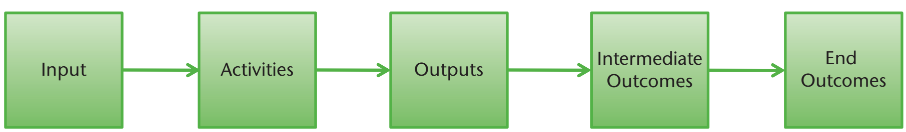


TABLE 13.1 Components of a Logic Model

Component	Definition and Evaluation Questions
Input	The resources that are being applied to the situation (e.g., funding for an after-school or tutoring program) or baseline data or starting point (fifth graders scoring poorly on the state math assessment the previous year).
Activities	Actions that occur as a result of the input. What occurs because of the input of these resources? (e.g., What activities occur because of the input that otherwise would not have been possible?) In addition to describing the activities, the evaluator documents the depth, breadth, and quality of the activities (e.g., Yes, the activities occurred, but what went on during these activities? Were they delivered in an appropriate and engaging fashion?)
Outputs of the activities	Changes in the participants immediately following their interaction with the activity. For example, a student engages in a math tutoring session. The tutor uses a number line to help teach adding and subtracting positive and negative numbers. The student grasps the concept easily and is excited. To document that the student has learned the concept, the tutor gives the student a quick assessment that is made up of five questions. The student gets all five correct. The tutor begins to plan the next strategy for their next session. Assessing the student immediately after the activity is the output.
Intermediate outcomes	Outcomes that develop when the participant continues to engage in the activity for a period of time but the outcomes are not the intended end results. In many cases they tend to be classroom-related outcomes and data tends to be teacher grade book data. For the fifth-grade student receiving the math tutoring, this might be an increase in math homework, improved behavior in class, participating in class more, and doing better on quizzes, unit tests, and any other classroom-based measures. Quarterly report card grades may also be examined as evidence for intermediate outcomes.
End outcomes	End results of the program and the primary reason the program was developed. In education, these outcomes tend to focus on changes in student performance on standardized measures such as annual state testing.

STEPS IN DESIGNING PROGRAM EVALUATION

The steps in conducting an evaluation are greatly influenced by the evaluator's approach (e.g., objective-based versus participatory). Presented following are some basic steps that an evaluator might consider when conducting a comprehensive evaluation of a program.

Step 1: Develop an Evaluation Plan

For most evaluation projects, a plan or blueprint is developed in the initial stages. An evaluation matrix is a common planning tool used by evaluators during this stage. Table 13.2 provides essential information for directing the evaluation process: evaluation objectives and or questions, identification of stakeholder groups from whom data will be collected, tools or methods used to collect data, type of data being collected (formative or summative), and timeline when data will be collected (e.g., pre-, post-, or follow-up measure).

Step 2: Develop Evaluation Capacity

After developing an evaluation matrix, evaluators work to develop what they refer to as **capacity**. This is the process used in developing the various evaluation tools needed to collect quantitative and qualitative data. Surveys, observational protocols, focus group and interview protocols are some examples of the tools that evaluators use to carry out their data collection activities. While the evaluator is typically the key individual responsible for developing these tools, participatory-oriented approaches "relinquish" control over instrument development and train stakeholders (e.g., parents, students, community members) to develop, collect, and analyze their own data and report their own findings (Flores, 2008).

Step 3: Collect and Analyze Evaluation Data

Following the evaluation matrix, the evaluator next collects the quantitative and qualitative data for each objective. After collecting the data, the evaluator analyzes the data using the appropriate methods. For quantitative data, the evaluator might use descriptive and inferential statistics. Identifying themes and patterns and other qualitative analyses are conducted on interview, focus group, and observational data. Triangulation of quantitative and qualitative data from different sources can help evaluators ensure they are collecting valid and reliable data.

Step 4: Report Writing and Dissemination

After collecting and analyzing the data, the evaluator writes a report and disseminates the findings to the client and/or stakeholder groups. In a successful evaluation, the evaluator has established a relationship with the clients and stakeholder groups involved in the process and the findings from the report will be used to make necessary changes to the program.

PROGRAM EVALUATION DATA AND CRITIQUING EVALUATION REPORTS

As you can see from the evaluation matrix in Table 13.2, data collected by evaluators can come from a wide variety of sources. Following is an example of school report card data from fifth graders involved in an after-school math tutoring project.

TABLE 13.2 Example of an Evaluation Matrix

Evaluation Objective	Stakeholder Group	Evaluation Tool Used	Type of Data	Timeline for Data Collection
To document depth, breadth, and quality of the after-school math tutoring program	Observation of the program with students and teachers	Observation protocol	Formative	Weekly observations conducted throughout the duration of the program
To document stakeholder perceptions of the after-school math tutoring program	Students Grades 6–8 Parents Teachers Grades 6–8	Student survey Parent survey Teacher survey	Summative Summative Summative	Post Post Post
To document student improvements in math	Students	School report card 1st and 4th quarters	Summative	Post
To document impact of the after-school tutoring program on student performance on state math assessment	Students	State math assessment (fifth grade)	Summative	Post

Evaluating Data from an After-School Math Tutoring Program

The after-school math tutoring program is a state-funded initiative. Data for the project are reported by the evaluator in a yearly progress report to the state. In this case, the state would like the evaluators to report the number of students who have improved their performance in mathematics by comparing their first-quarter grades in math to those of the fourth quarter. Analyze the data in Table 13.3, and based on your analysis, report your findings for this evaluation component in Exhibit 13.1.

Critiquing Evaluation Reports

When evaluating program evaluations, it is important to do so within the approach or framework used to guide the evaluation. For example, it would be inappropriate

TABLE 13.3 Data from School Records for Fifth-Grade Students in Math

Student	First-Quarter Math	Fourth-Quarter Math
1	C	B−
2	C	C
3	D	C
4	B	C
5	C	C
6	B	B
7	D	C
8	D	F
9	B	A
10	B+	A−
11	B	B
12	B	C
13	B	B
14	B+	A
15	B+	C
16	C	A
17	B−	B+
18	C	B−

EXHIBIT 13.1
Math Outcomes for State Annual Report

Criteria	Outcome
Number of students who increased math performance by half a grade or more	
Number of students who decreased math performance by half a grade or more	
Number of students who neither increased nor decreased math performance	
Number of students who neither increased nor decreased math performance because their first-quarter grade was already at its maximum	

to evaluate a goal-free evaluation report using criteria for an objective-based approach. Evaluators work continually to improve their methods for collecting and analyzing data but judge the success and rigor of their own work by the change their evaluation reports make. Do the clients use the evaluation findings to make notable improvements in the program? Are those who participate in the program now better served as a result of the findings the evaluation brought to light? As stated earlier, evaluation reports that sit on the shelf and do little to change the program are considered to be a failure for professional evaluators, particularly those who value utilization of findings as much as the data collection and analysis process.

TRAINING AND CAREERS IN PROGRAM EVALUATION

Many students wonder how evaluators get involved in program evaluation, and where they receive their training. These are both good questions. While program evaluation today is certainly a much more recognized field than it was thirty years ago, it is made up of a wide range of professionals. Some of these professionals have received specific training and course work in program evaluation, while others have learned much of what they know from on-the-job experiences. No specialized degree or certification is required for the title of evaluator. There are a number of colleges and universities that offer coursework as well as advanced degrees in program evaluation. While coursework varies by institution, most focus on quantitative and qualitative methods, program evaluation theory, and ethics; a degree usually requires practicum experience as well.

As in any field, program evaluators come from a wide range of backgrounds and experiences as well as from different philosophical and methodological perspectives. Many times faculty at colleges and universities serve as program evaluation consultants working with area school districts, agencies, for-profit and not-for-profit programs, and other institutions of higher education. Private evaluation consulting companies also exist. Companies such as these hire program evaluators. Public agencies at both the state and federal level hire program evaluators in full-time positions to conduct internal evaluations within a setting as well as to conduct single- and multisite evaluations.

The American Evaluation Association is an international organization devoted to improving evaluation practices and methods, increasing their use, promoting evaluation as a profession, and supporting evaluation to generate theory and knowledge. This organization has approximately 4,000 members and representatives from 50 states and 60 foreign countries. Each year the association hosts an annual conference in the United States that focuses on a theme such as collaboration, methodology, or utilization (see http://www.eval.org/News/news.htm).The association is also comprised of special interest groups (SIGS) that specialize in certain areas or topics (e.g., teaching program evaluation or environmental evaluation.

SUMMARY

Although program evaluation uses the same quantitative, qualitative, and mixed-methods approaches as does applied research, program evaluation is typically used for decision-making purposes. One main difference between program evaluation and research is in how each accesses the settings and works with participants. Whereas researchers might look at subgroups or individuals in a particular setting, program evaluators are generally hired or contracted by a group to judge the worth or merit of a particular program as a whole.

Program evaluators rely on summative and formative evaluation. Summative evaluation focuses on gathering specific kinds of outcome data, such as test scores and final results, to determine whether the project met its overall goals. Summative data is provided to the client or group that commissioned the evaluation work at the end of the project. Formative data is collected by the evaluator as the program is occurring and is used to modify or refine program activities to improve program quality before the program has been completed. Because of the speed at which formative evaluation needs to be presented back to the client, formative findings can be presented in oral reports, memorandums, or conference calls between the evaluator and the client. Summative findings are typically presented in a more formal evaluation report at the end of the project.

Internal evaluators are persons who are employed by or associated with the group that is having an evaluation conducted. An external evaluator is someone who is hired from outside the group to conduct an evaluation. Since internal evaluators are already associated with the group that is conducting the evaluation, they are aware of how the group functions, its processes, language, and politics. Since they are not a part of these systems, many times external evaluators have to work to establish trust with the group that they are evaluating.

There are several different evaluation approaches that evaluators utilize. Objective-based, goal-free, expertise-oriented, and participatory-oriented evaluations are examples of different approaches, each geared to focus on different aspects of the evaluation, each with different intentions driving their methodologies.

KEY CONCEPTS

benchmarks
client
expertise-oriented evaluation
external evaluators
formative
goal-free evaluation
internal evaluator
logic model
objective-based approach
participatory-oriented evaluation
summative

DISCUSSION QUESTIONS AND ACTIVITIES

1. An external program evaluator has been hired to examine the science program at a suburban high school and to show how the program affects students' future success in college. The high school, however, already has a staff member who serves as an "informal" internal evaluator collecting data on student performance. Discuss what expertise these two evaluators can bring to the project and what data possibilities each can collect.
2. A program evaluator is working on a project and is having difficulty building trust with the client. The client fears that the evaluator is only interesting in "digging" up bad stuff about them and the things that they are not doing with their program. Give an overview of the differences between formative and summative evaluation findings and then discuss how the evaluator might go about using these two approaches to improve the rapport with the client.

3. Discuss some possible issues that program evaluators might face on a regular basis, particularly because of the close working relationship that exists between them and their clients.

SAMPLE PROGRAM EVALUATIONS

Converse, N., & Lignugaris/Kraft, B. (2009). Evaluation of a school-based mentoring program for at-risk middle school youth. *Remedial and Special Education, 30*(1), 33–46.

Lewis, N. S. (2006). An evaluation of a master's degree in K–8 mathematics and science: Classroom practice. *School Science and Mathematics, 106*(6), 231.

Simpson, J. S., & Parsons, E. C. (2009). African American perspectives and informal science educational experiences. *Science Education, 93*(2), 293–321.

SUGGESTED READINGS

Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Education Evaluation and Policy Analysis, 14*(4), 397–418.

Flores, K. S. (2008). *Youth participatory evaluation: Strategies for engaging young people.* San Francisco: Jossey-Bass/Wiley.

Frechtling, J. A. (2007). *Logic modeling methods in program evaluation*. San Francisco: Jossey-Bass/Wiley.

MacNeil, C. (2002). Evaluator as steward of citizen deliberation. *American Journal of Evaluation, 23*(2), 45–54.

Patton, M .Q. (2002). *Qualitative research and evaluation methods.* Thousand Oaks, CA: Sage.

CHAPTER FOURTEEN

IDENTIFYING AND RESEARCHING A TOPIC

CHAPTER OBJECTIVES

- Summarize the steps for identifying a topic and developing a focused research question and statement of purpose
- Distinguish among the types of sources that can be used in a research paper
- Be able to identify sources as primary or secondary sources
- Distinguish between empirical research reports and nonempirical reports, such as opinion pieces and how-to or experiential reports
- Develop competence in finding research studies on a topic using several different strategies, such as locating a review of literature, searching databases, following citations, using people as sources, and finding and evaluating Internet sources
- Develop skills in identifying and summarizing key information from scholarly sources

GETTING STARTED

Are you feeling a little overwhelmed with all the new concepts and strategies that are involved in studying educational research? Well, you are not alone! One of us (D.T.S.) has described student experiences in taking a course in educational research as similar to learning to drive at 200 miles per hour. Especially when students begin to read studies published in professional journals, they often feel that there are just too many strange new concepts, too much information, and too many different types of research to expect anyone to navigate successfully without getting lost. This chapter presents some ideas to help you in your travels through the professional literature. We cannot tell you everything you need to know before you "hit the road." But we can give you some tools to help you select and define a topic, find information on your research topic from appropriate scholarly sources, and identify and summarize key information from these sources. We start at the beginning by considering how to select or generate a research topic.

IDENTIFY A RESEARCH TOPIC

The first task in developing a research study is coming to some decision about a research topic. For some students, this can be an exciting process: finally pursuing a topic that you have been interested in all your life but have never had the time or the opportunity to pursue. (Not that you have a whole lot of time now that you are in graduate school, but at least you can't say you don't have the opportunity.) For other students, coming up with a topic can be a horrible, painful experience: torn among several subjects, one more interesting than the next, wanting to explore everything, leaving no stone unturned, and believing that it might be this study (or maybe the one right around the corner) that will somehow ultimately change the world . . . if you could only wait one more day before having to decide. And then there are the students, many of whom are new to research, who are keenly interested in a certain broad topic but know little about it or the current research supporting it and have no idea how to begin a study on it. Most graduate students in an introductory course to educational research are in this last group. So relax, take a deep breath, and let it out slowly. This is exactly where you should be right now in this process, and it is a wonderful place to be: ready for adventure and exploration.

Most students in a research course begin by deciding on a couple of broad research topics. In educational research, these topics can be wide ranging, including but by no means limited to the following: emergent literacy, technology integration, strategies in educational leadership, or behavioral intervention practices.

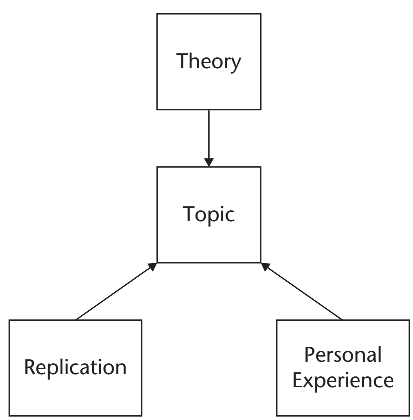
Many students have the common misperception about educational research that it must take place within the context of a school building or classroom. This is a limited view of educational research, and although most educational research does indeed take place more or less within such a setting, many studies are done in nontraditional educational settings outside the brick-and-mortar infrastructure of the school building. For example, research conducted on how companies train new employees using distance learning and other technological hardware and software would certainly be considered in the realm of education-related topics and suitable for study in an educational research course. Educational research also includes research on programs and activities in after-school settings and in the communities surrounding schools.

Although there is no specific formula or proven method for selecting a topic to pursue, students and most professional researchers gravitate toward a specific topic for one of three reasons: past experience, theory testing, or replication of previous research (see Figure 14.1).

Experience

For most students in a research course, past experience is one thing that drives their interest in pursuing a particular topic. Past experience can be a strong motivational factor for professional researchers, who sometimes devote their whole lives to pursuing an area of study. For those in a graduate course on educational research, experience may be related to their current careers as teachers in the

FIGURE 14.1 Ways to Generate a Research Topic



classroom, school psychologists, counselors, or as school administrators. In those situations, individuals should ask themselves questions such as:

- I wonder if we did it this way, rather than the old way, would that make a difference?
- I wonder if I taught the class using an activity first, followed by an explanation of the content, would that make a difference in my students' understanding and increase their performance on the next chapter test?
- I wonder what intervention would be best for a student with multiple disabilities?
- I wonder what that new teacher in fifth grade is doing to keep her students so attentive during class?

All of these questions are cornerstones to developing a more comprehensive, researchable topic. Keep in mind that researchers do not always pursue a topic of interest because they had a positive experience with it in the past. In fact, many researchers pursue topics based on past negative experiences for the sole purpose of improving practices and making them better for the next generation of learners. For example, one of our students explored student experiences in gifted education programs in her proposal because she had experienced significant problems within her own high school gifted education program.

Theory

Research studies can also be developed based on theory, as discussed in chapter 1. For some researchers, a single theory might provide the focus for most of their research. For example, Robbie Case (1992) conducted many studies of Piagetian theory. These studies eventually led him to construct his own theory that blends Piaget's original stages with concepts from information-processing theory. Students might use theories presented in any of their classes as the basis for a research study. For example, several of our students have examined implications of a well-known theory of parenting styles (Baumrind, 1967) in relation to educational issues such as student achievement, discipline problems, and attention disorders.

Replication

The third approach to doing a study is through **replication.** Although this approach may sound complicated, in essence, it is a "do over but do better" approach. Typically, a researcher conducts a replication by first selecting a research study that she or he recognizes (or learns through others' reviews of studies) as flawed; that is, it either uses poor methods to collect the data or sample participants or is poorly designed in other ways. Some of these issues may be directly related to poor planning and practice by the researcher. Aware of these methodological limitations, the researcher conducts the study again, paying careful attention that such oversights do not repeat themselves. Other times, an existing study is replicated not so much because of methodological flaws but because the researcher wants to redo the study with a slightly different population (for example, using fifth graders instead of fourth graders) to see if the findings are the same. For example, one of our students replicated a classic study of the benefits of preschool programs, but he included measures of attitudes toward school as well as student achievement and extended the follow-up for more years. Some other sample research topics from our students' proposals that were based on experience, theory, or replication are listed in Box 14.1.

REFINE YOUR TOPIC AS YOU SEARCH

After a topic has been selected, the next step in the process is an initial search of the literature. This initial search may consist of finding and reading only a handful of research studies, literature reviews, and other articles on the topic. You are in the "exploratory" stage of your study, where your goal is to simply increase your knowledge on your selected topic. As you read, you should pay close attention to the issues and questions that are addressed within your topic. For example,

BOX 14.1
Sample Topics for Research Proposals

- Student perceptions of school violence
- Approaches to classroom management at the middle school level
- Involvement of single parents in their children's education
- Development of new instructional strategies for students with reading problems
- Testing the effectiveness of an intervention to build social skills in a child with autism
- Attention deficit hyperactivity and strategies to improve behavior
- The principal's role in training new teachers
- Improving one's own instruction in English Language Arts

you select a topic that involves successful urban schools and teachers. Certainly a broad topic! You might begin your investigative process by selecting a few articles that focus on successful urban schools and teachers and find that the following issues are identified:

- Instructional techniques related to teaching diverse populations of urban students
- Characteristics of successful urban schools
- Characteristics of quality urban teachers
- Training quality urban teachers

Your initial search has identified some of the major areas of investigation around the topic of successful urban teachers and students, which should help you to think about some research questions that could be explored on your topic.

Generally, research studies stem from a research question (or in some cases, a series of questions). One mistake that many students make when selecting a topic is to think that they need to generate a final, detailed research question at the exact moment they come up with their topics. In general, we recommend that students start with a general question that can be further developed after they begin to read the literature related to it. Students should not avoid a topic just because they cannot think of a research question to ask. Sophisticated research studies can begin with simple questions that are refined repeatedly as you review past research. Because the research question is the seed from which the study will eventually grow, it is imperative that the question be what is often referred to as *researchable* or *doable*. **Researchable questions** can be answered through the systematic collection of data and clearly meet ethical guidelines. (The issues of ethics were discussed in detail in chapter 1.) Remember that all human research should protect participants from harm, provide confidentiality, and include informed consent. Researchable questions should also clearly define the variables (quantitative research) and identify the group or setting (qualitative research) being investigated.

As discussed in chapter 2, researchers using experimental, causal-comparative, and correlational approaches all measure or manipulate variables. Research questions for these types of quantitative research are often stated in terms of variables. Let us say that you are interested in tutoring in reading and its influence on reading performance. Your initial search of the literature might identify several variables related to reading performance, such as a parent's reading to a child, instructional methods used by the teacher, or gender. This search might also provide examples of how the variables were operationally defined, measured, or manipulated. If you decide to pursue an experimental or causal-comparative approach, your research question will identify the independent and dependent variables. Later in the proposal, you would design a study that would control for the effects of the extraneous variables identified in your review of the literature. If you decide to do a correlational study, your research question will identify the variables that you expect to be related.

Unlike their quantitative counterparts, qualitative researchers begin with a research question that does not identify specific variables. Instead, they identify a group or setting that they wish to study in depth. Qualitative researchers are concerned with meaning, which essentially is the unique way people make sense of their lives. Meaning is always described from the perspective of the participants—those who are being studied. For example:

- How would parents describe their experiences in an inclusion preschool?
- What resources do they assume will be available for their children?
- How welcoming is the school to parental involvement?

Additionally, qualitative researchers are concerned with processes (descriptions of how education is practiced) rather than outcomes (end results of those practices). Qualitative researchers are likely to focus on processes in their research questions. For example, a quantitative study of an after-school program might have reading performance as its outcome. A qualitative researcher would be more interested in the processes that were used during the tutoring program:

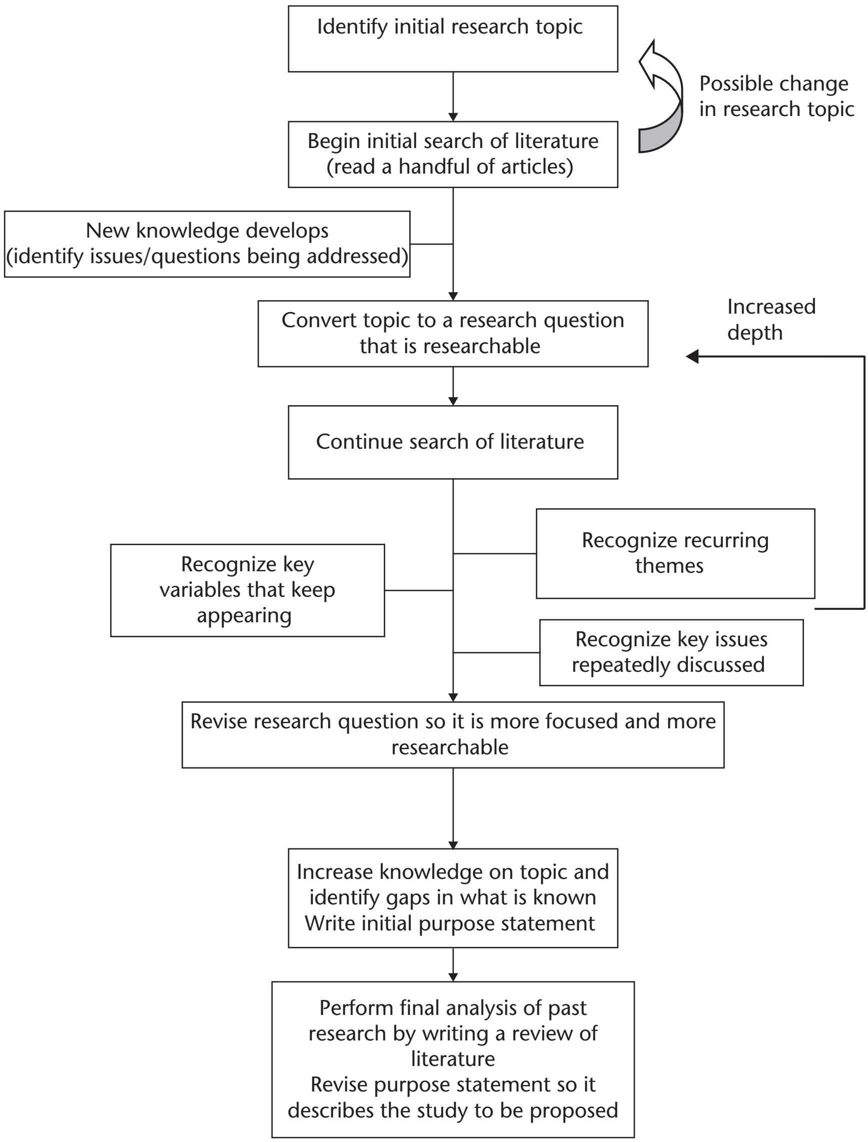
- What materials were used by the instructors?
- How did the students and instructors interact during the tutoring sessions?
- How did the students feel about their participation in the tutoring sessions?

When data from qualitative studies are analyzed, the researchers abstract and describe general themes—major recurring issues or concepts that they use to summarize their interpretations of their data. Themes from previous studies may help a beginning researcher develop a focus for his or her research question.

As you begin to explore the research on your topic, make note of the variables, themes, processes, and meanings that are reported in the studies. You will likely include some of these in your research question.

As you complete your initial search of the literature, you should develop new knowledge on the topic, including the themes or issues that reoccur, the settings or groups that have been studied, and the results of previous studies. This knowledge should help you to ask a more focused and researchable question and to begin to generate ideas about a possible study that could be proposed. The research question will be continually refined down to a much more specific question and a statement of the purpose of the study as you complete the search of the literature. See Figure 14.2 for a summary of this process.

FIGURE 14.2 Development of a Research Question



Looks complicated, doesn't it? However, remember that it will take several weeks to arrive at your final research question and statement of purpose. All you need to begin your search of literature is a topic to explore!

SEARCH THE LITERATURE

Every topic and every literature search is different, but all involve a substantial commitment of time and energy. In an age of expanding electronic resources, many of the tools used in searching the literature are computer based and digital. However, we urge all students to plan to spend a substantial amount of time in their college libraries, for two reasons. First, reference librarians are experts in information literacy and are often the most knowledgeable technologists on college campuses. They can guide students in the use of ever-changing electronic resources and are often familiar with the terms used in databases of professional literature. Most libraries offer one-on-one consultations for students conducting literature searches, and the library staff are available at the reference desk at hours that far exceed the office hours of all three of the authors of this book combined! So use the librarians whenever you have a problem, but especially early on in your search. Second, many of the sources needed in a research proposal are not online. So if you try to conduct a literature search using only online sources, you will miss critical information.

As you search the literature, it is critical to record citation information from each source. **Citation information** includes the names of authors, date of publication, title of work, and information about where you found the work (e.g., journal name, volume, page number; book title, publisher; city; information about and location of electronic source). The way each source is cited in the text and the reference format of the paper are described in the *Manual of the American Psychological Association*. For complete information consult the most recent edition of the manual.

Types of Sources

Whether sources are online or print based, there are many different types of sources that can be used in a review of literature, including the following:

- Articles in professional journals
- Government and technical reports
- Conference proceedings and papers
- Reference books

- Monographs
- Books (general circulation)
- Master's and doctoral theses
- Web sites
- Magazines and newspapers
- Personal communications: information from interviews, presentations, lectures

A good researcher knows when and how to use each type of source effectively. Sources for a research proposal are considered strongest if they are **peer reviewed**, meaning that a panel of researchers (considered your peers!) has reviewed the articles and rated the quality of the methods used in them. Articles in professional journals and some conference papers are the sources most likely to be peer reviewed. However, all of the sources listed may be useful in some way, and not all articles in professional journals or conference proceedings are peer reviewed.

Another criterion used to evaluate sources in a review of literature is whether the sources are primary sources that describe empirical research. A **primary source** is an article that describes original research conducted by the author of the article. By contrast, a **secondary source** is any article written by someone who is describing research done by others, so the description is secondhand. **Empirical research** is a term applied to studies in which a research question has been examined by systematically collecting and analyzing data. These articles usually (but not always) include a section labeled as the method section, in which are described the persons or groups participating in the study and the methods used to study them. A secondary source, such as a book, may describe empirical research, but it usually gives less information on the research methods and is usually considered to be a weaker type of source. Professional journals include both primary and secondary sources as well as empirical and nonempirical sources. One useful type of secondary source is a **review of literature** in which the author provides a comprehensive overview and critique of prior research studies on a given topic. Reviews of literature are useful sources when students are beginning their search of literature; strategies for finding and using reviews of the literature are discussed following. Most empirical research articles provide a brief review of literature before their method sections, and articles that are solely reviews of literature have more comprehensive references but do not include a method section. **Theoretical articles** also provide overviews of empirical research, although the studies reviewed are selected based on their relevance to the theory examined.

Nonempirical articles frequently found in journals include **opinion pieces** or commentaries in which experts discuss their ideas or conclusions based on previous research, and "**how-to**" or **experiential reports** in which practitioners discuss how to implement an educational practice based on their own experiences. Whereas opinion pieces can be useful in identifying issues of debate in an area, the quality of the information depends on the credibility of the expert. Experiential reports may provide creative ideas for educational practice, but because there is no systematic collection of data and results are based on the experience of just one practitioner, they do not provide strong evidence.

Often the first thing you read in any of these sources is an **abstract,** which is a summary of the article or study. It may be difficult to tell if a source will be an empirical or nonempirical source based on just the abstract. Nonempirical sources are more likely to discuss in the abstract the general issues related to a topic, whereas empirical sources are more likely to describe the groups studied and the methods used in the study (for example, surveys, observations). Exhibit 14.1 provides a comparison of sample abstracts for an experiential report and a primary, empirical study.

EXHIBIT 14.1
Distinguishing Among Types of Research: Experiential Report Versus Empirical Study

Empirical Study

Praisner, C. (2003). Attitudes of elementary school principals toward the inclusion of students with disabilities. *Exceptional Children, 69*, 135–145.

Abstract: <u>A survey of 408 elementary school principals</u> found about 1 in 5 had positive attitudes toward inclusion while most were uncertain. Positive experiences with students with disabilities and exposure to special education concepts were associated with positive attitudes toward inclusion and more positive attitudes were related to less restrictive placements. (Retrieved April 18, 2005, from ERIC database.)

Indicates measure used and size of sample.

Experiential Report

Tooms, A. K. (2003). Bring in Mac. *School Administrator, 60*, 22–25.

Abstract: Describes how an <u>executive coach</u> nicknamed Mac helped a beginning principal develop her leadership skills. <u>Also provides list of questions to ask oneself</u> before hiring an executive coach. (Retrieved April 10, 2005, from ERIC database.)

Based on one person's experience. Goal is to provide practical advice. No mention of methods of data collection.

Books, monographs, and reference books may also provide overviews of research, although these are likely to be several years older than comparable reviews from similarly dated journal articles. For some topics, government and technical reports may be useful sources, especially if the agencies or persons involved are professional researchers. Web sites provide sources that can vary enormously in quality. Remember that anyone anywhere can put information on the Web, so information from the Internet should be cited with caution. Many Web sites for governmental agencies, educational institutions, and professional organizations provide useful information, but one must always carefully evaluate any information obtained from a Web site for potential bias and for accuracy. The types of Web sites that are especially useful to students in educational research are discussed in Box 14.2. More detailed discussion of using and evaluating Web site information is discussed later in this chapter.

BOX 14.2
Useful Web Sites for Finding Information and Sources for Reviewing the Literature

Web sites of governmental organizations, professional associations, and research organizations often provide resources for both practitioners and researchers. Look for links that are labeled *research*, *publications*, or specific topics. The following are some Web sites that may be useful.

Governmental Organizations

Centers for Disease Control and Prevention (CDC):

http://www.cdc.gov

National Center for Education Statistics (NCES):

http://nces.ed.gov

U.S. Census: American Factfinder:

http://factfinder.census.gov

U.S. Department of Education:

http://www.ed.gov/index.jhtml

Strategies for Searching the Literature

Finding sources typically entails the use of several different strategies. The strategy that works best will depend in part on how focused your topic is and whether there has been extensive previous research on it. So select a couple of strategies from those described below that you think will work for your topic.

Browse Key Journals Many students begin with a general topic, such as learning disabilities or child neglect and abuse, and have trouble initially thinking about how to focus or narrow it down. For some topics, it may be helpful to simply browse through a journal that focuses just on that topic. For example, the *Journal of Child Abuse and Neglect* and the *Journal of Learning Disabilities* are journals dedicated to those topics. If you are interested in these areas, you could read through the

Professional Associations

American Association of Colleges of Teacher Education (AACTE):
http://www.aacte.org
American Educational Research Association (AERA):
http://www.aera.net
American Psychological Association (APA):
http://www.apa.org
Association for Qualitative Research:
http://www.latrobe.edu.au/aqr/
Council for Exceptional Children (CEC):
http://cec.sped.org
National Association of School Psychologists (NASP):
http://www.nasponline.net
National Association of Special Education Teachers (NASET):
http://www.naset.org

Research Organization

Action Research Special Interest Group (SIG) of American Educational Research Association (AERA):
http://www.coewestga.edu/arsig

abstracts of the studies in one of these journals and select a couple of articles to read. Reading articles in the most recent issues of a journal might provide ideas about an issue or problem that could form the focus of a research proposal. It is also a quick way to locate specific articles on your topic and to identify some of the key researchers on a topic. We suggest that you browse through a couple of years of journal issues until you feel that you have a clearer idea of the topic that you want to research. Then it is time to move on to another strategy, because it is unlikely that a single journal will provide all of the current information on a topic.

Find a Review of Literature Although your ultimate goal is to find primary sources on your topic, many students find it useful to begin their literature search with a review of literature. As noted earlier, this is a secondary source, but a good review of the literature will help acquaint you with the current issues, authors, research methods, and problems for a particular topic. The good news is that it does a lot of your work for you! The bad news is that current reviews of literature simply do not exist for all topics. One approach to finding a review of literature is to use the words *review of literature* as one of your terms in searching databases, a topic discussed next.

Search an Electronic Database A database is a collection of information on articles and materials that have been published, presented at conferences, or created by various educational groups or individuals. A database includes information about an article, such as the author, title, number of pages, where it was produced or published, and an abstract or summary of the contents of the article. Most databases allow you to combine several terms to find articles. The advantage of computerized databases is that searching with a specific set of terms is fast. The disadvantage is that computers look only for the words that you input. So you need to think about how to best describe the information that you want. To do this, begin with your research question.

As noted previously, your research question should begin as a broad statement of a topic that interests you. We suggest that you write out your research question and describe in a couple of sentences what you would like to find out. You next identify some words or phrases that capture the most important issues or elements of your topic. These may be variables such as a particular instructional approach (for example, problem-based learning, phonological training), characteristics of a person or group (for example, learning disability, gender, rural), or a type of setting (for example, multiage classroom, after-school program, high school). Write down synonyms or related words so you can try out different searches. Once you have a list of terms that capture your topic, select a database and open the search tool. The art of database searching involves learning how to input terms that will connect with the articles most related to your topic. Each database has been organized using a set of controlled vocabulary, which is a group of terms or keywords that the database creators use to categorize the articles in the database. Your next step is to select keywords that are most closely related to the terms you have used to describe your topic. ERIC (Educational Resources Information Center) and PsycINFO offer a tool called the Thesaurus that describes the keywords used and related terms. You can select several keywords and enter these into a search box. Finding the right keywords may involve trial and error. We suggest that you try out several combinations of keywords and see which ones seem to yield the best sources. You can also use an author's name as a search term, a strategy that is often productive because authors frequently conduct multiple studies on a given topic. Also check with your instructor and librarians for new tools or strategies included in the latest versions of the databases to make searching easier. The most commonly used databases in education are listed in Table 14.1, along with a description of what these databases include.

Terms can be combined in a search using words such as *and*, *or*, and *not* that are referred to as "Boolean operators." Inserting *and* between the keywords instructs the database that you want to find only articles that contain both keywords. Inserting *or* between the keywords instructs the database that you want to find articles that contain either of the keywords but do not have to contain both keywords. Using *and* will narrow your search to fewer articles, and using *or* will

TABLE 14.1 Commonly Used Databases in Education

Database Name	What It Includes	Additional Features
ERIC	Journal articles and documents sent by their author directly to ERIC. ERIC documents include papers presented at conferences, master's theses, doctoral dissertations, government reports, curricula, books, chapters in books, and other materials.	Most ERIC documents are available full text. Journal articles may be available full text.
PsycINFO	Journal articles and books on topics in psychology and related fields, including education.	PsycINFO provides bibliographies with article abstract; allows following of article citations in later publications.
Education Full Text	Journal articles, monographs, and yearbooks related to education.	Substantial overlap with ERIC; however, it does cover 40 journals not indexed in ERIC.

Note: ERIC = Educational Resources Information Center.

broaden it. The operator *not* indicates that articles containing that keyword should be eliminated from the search. This may be useful if your first search turns up articles on settings or groups that do not interest you, such as college students or corporate training centers.

ERIC and PsycINFO provide ways to refine or limit a search to only research reports. PsycINFO allows you to limit the search to reviews of literature or specific types of research studies such as case studies or experimental studies. Because the database formats change frequently, you should check with your instructor and librarian for more recent information about how to refine your search. Even if the database does not include a way to limit the search you wish to do, you can use words describing what you want as keywords. Again, the art of database searching involves learning how to use all of these tools to input terms that will connect with the articles most related to your topic.

The initial output from a database search will be a list of citations for the articles that match the terms used in your search organized by date. The citation will include the authors' names, title of the article, and place of publication or presentation. In ERIC there will be either an ED or EJ number included with the article. EJ indicates that the article was published in a journal, and ED indicates that it is an ERIC document. ERIC documents include materials that have been submitted to ERIC by educators in a variety of settings. The documents include papers presented at conferences, organizational reports, curricula, manuals, and governmental reports. You will need ED numbers for the reference citations for most ERIC documents, so if you plan to use these articles, jot down the ED numbers. You do not need to include EJ numbers in references. However, some electronic sources may include a number called a Digital Object Identifier (DOI). The DOI is a number representing a new system for managing electronic sources. It usually includes the number 10 followed by a string of numbers. Consult the Publication Manual of the APA for further information. Read through the list of titles and click on the links for the articles that sound most related to your topic. A bibliographic entry and abstract for the article will appear.

Usually the abstract contains the information most useful in deciding if you want to obtain and read the article. Remember that your major goal is to find empirical sources. There is no foolproof way to determine if a source is empirical based on the abstract, so look for descriptions of the research methods and results in the abstract. If you find several articles that are related to your topic, you may want to try out the strategy of following citations described later in the chapter.

Sources found through databases may include books, journal articles, organizational reports, curricula, conference papers or proceedings, or other materials. Some sources may be available as **full text**, which means that the entire text of the source can be accessed through the computer. In some cases, the source appears as an html (hypertext markup language) file with sidebar links to sections of the article. At other times, you will need to click a .pdf link (portable document format), and a photocopy of the original article will appear. We strongly caution our students against limiting their searches to full-text articles because this will eliminate access to important sources.

Follow the Citations Sometimes students find a couple of related studies but not much more on their topics. Following citations is one way of expanding the research base in this case. PsycINFO provides a link to more recent studies that have cited the study that is presented. Citing means that other studies have listed this study in their references, and therefore they are likely to discuss similar issues or problems. Clicking on the link brings up a list of the studies that have cited the one produced by your search. Social Science Citations is a database that allows you to enter any reference and find more recent studies that have cited that reference.

Another strategy for following citations is to examine the references in the studies you have found and find those that sound most related to your topic. You can enter the author's name and title of the article into a database search to see if the article is available online or simply locate the journal or book in the library.

Consider People as Sources We often remind our students that people can be sources of information, too, and can even be cited as sources in a proposal. Consider first the people you know who have expertise in the area that you wish to research. These may include persons at your workplace, school, or college and people in the community. Documents, training materials, class handouts, and interviews with people can all serve as sources of information on your topic. One advantage of local people is that they may be able to provide information that is specific to the setting or group that you want to research. One of our students obtained a wealth of information for a paper on homeschooling from an interview that she conducted with a professor who was homeschooling his own children. Another obtained extensive data on the number of students attending college following an interview with her high school's guidance counselor.

The Internet also provides a way for students to easily contact authors and researchers by e-mail with questions about their research. Our students have found that most researchers are happy to help students with their research as long as the students have done their initial homework. You would most likely not get a reply from Howard Gardner if you wrote to simply ask what he knows about multiple intelligences. However, our students have found that researchers graciously respond to their requests for specific data collection tools and answer targeted questions about their methods that were not discussed in their articles. (One of our students received such a response from Gardner after she wrote a carefully thought-out e-mail to him.) It is exciting to find that the people in the databases really exist and like to help out beginning researchers. When communications with people cannot be accessed by others, they are referred to as **personal communications** and are cited in the proposal, along with information on the name of the person and the date when the communication occurred, according to style rules in the *Publication Manual of the American Psychological Association* (APA). For example, if the information in this sentence was based on a conversation with Marguerite G. Lodico on March 7, 2005, you would cite this parenthetically as follows: (M. G. Lodico, personal communication, March 7, 2005).

E-mail discussion groups or listservs are another source of information for research proposals. A **listserv** is a group discussion of a topic of common interest to which one subscribes through e-mail. Messages from all members of the listserv are sent to your e-mail address, and you can also send questions or messages to the listserv discussion. Typically, listserv discussions involve threaded conversations in which multiple people comment on an issue or on each other's comments by sending messages to the listserv. You can find the names of listservs related to education and instructions on how to subscribe and unsubscribe at http://www.edwebproject.org/lists.html.

Search the Internet for Scholarly Sources As noted earlier, the Internet is a complex network of information, both good and bad. We advise extreme caution in using any information from a Web site. If you cannot determine the author of information obtained from the Web or the date it was produced, you should not use it in a research proposal. You'll likely find research done on a topic that is available only through an organization or governmental agency that publishes this information on the Internet. Many researchers provide information on their work on their Web sites. So to use an analogy from skiing, using the Internet as a source is the advanced slope; you need all your skills operating at full capacity to navigate among the reputable and untrustworthy sources available that also change frequently. Table 14.2 presents search engines useful to educational researchers.

Next we discuss issues that you should consider when evaluating the quality of information obtained from Web sites.

Authority It should be clear who produced or is sponsoring the site, and verifiable information about the sponsor should be provided. Look for a link to the information about the author, organization, or sponsor (often called "About Us") and links that allow you to send messages to the Web site creator. A good way to see if an organization is legitimate is to look it up in the *Encyclopedia of Associations*, a book available in many college library reference sections. Sometimes you may need to return to the home page of the Web site by clicking the link to "Home" or truncating the URL (uniform resource locator) back to the domain (the letters

TABLE 14.2 General and Academically Focused Search Engines and Their Web Sites

	Broad Search Engines		Academically Focused Search Engines		
	Google www.google.com	**Yahoo! Search** search.yahoo.com	**Library Index** www.lii.org	**Infomine** infomine.ucr.edu	**Google Scholar** scholar.google.com
Size (*Note:* Estimates of size vary widely)	Claims to be most comprehensive but has dropped its index count so there is no way to estimate	More than 20 billion total "web objects" Index directories available	Subject directory of more than 20,000 sites created by public librarians	University of California librarians created this directory of 120,000 pages	Directory of Web sites with scholarly literature such as peer-reviewed papers, theses, books, preprints, abstracts, and technical reports
Special features	Additional databases including Book Search, Blog Search, Patents, and Images	Shortcut links to dictionary, synonyms, and encyclopedia	Links to dictionaries and encyclopedias Good source for organizations with research resources	General Reference link gives access to dictionaries, style guides, maps, news, and organizations Includes links to databases, electronic journals, electronic books, bulletin boards, mailing lists, online library card catalogs, articles, and directories of researchers	Search automatically analyzes and extracts citations to other sources and presents them as separate results

(*Continued*)

TABLE 14.2 (*Continued*)

Broad Search Engines			**Academically Focused Search Engines**		
	Google www.google.com	**Yahoo! Search** search.yahoo .com	**Library Index** www.lii.org	**Infomine** infomine.ucr .edu	**Google Scholar** scholar.google .com
Boolean Logic	AND assumed between words Capitalize OR - excludes a word + allows retrieval of stop words (for example, IN, AND)	Accepts AND, OR, NOT, or AND NOT Words must be capitalized to serve as operators	AND implied between words Accepts AND, OR, NOT, and ( ) for nesting	AND implied between words Accepts AND, OR, NOT, and ( ) for nesting	AND assumed between words Capitalize OR - excludes a word + allows retrieval of stop words (for example, IN, AND)
Phrase Searching	Put " " around words to search for phrase	Put " " around words to search for phrase	Put " " around words to search for phrase	Put " " around words to search for phrase	Put " " around words to search for phrase
Advanced Search	Allows you to search for pages that have all your search terms, contain an exact phrase, have been updated recently	Allows you to search for pages that have all your search terms, contain an exact phrase, have been updated recently, or are within a certain domain (for example, .edu or .org)	Allows you to search using author, title, subject, publisher, keyword, or sponsor	Allows you to search using author, title, subject, keyword, research discipline, or Library of Congress subject heading	Allows you to search for pages that have all your search terms or contain an exact phrase Allows search using author, title, subject, publication, or date
Link to Help for Searching	http://www .google.com/intl/ en/about.html	http://help.yahoo .com/l/us/yahoo/ search/	http://lii.org/pub/ htdocs/search_ help.htm	http://infomine. ucr.edu/help/	http://scholar. google.com/intl/ en/scholar/help .html

following the first period in the URL). If you cannot determine who produced the Web site, you should not use it as a source because its authority cannot be established.

Accuracy Expect few, if any, grammatical or spelling errors in a credible Web site. Sources of information should be clearly identified. Factual information such as dates and names should match those found in almanacs, encyclopedias, or other reference books. If the information on the Web page contradicts information in reputable sources, the author should say so and explain why. If the accuracy of a Web site seems dubious, do not use it as a source.

Objectivity Look for a statement of the sponsor's point of view, and be sensitive to any expressions of bias. Ask yourself, "Where is the author coming from?" Take the author's point of view into account when you use the material in your paper. An unbiased Web site will include links to other Web sites or to authors providing contrasting points of view. If the Web site provides no balance or never acknowledges other points of view, use the information with extreme caution. If there is advertising on the Web site, it should be separate from the information content. Consider whether the advertising is related to the content of the Web site. An advertisement for a company producing medications for attention-deficit hyperactivity disorder (ADHD) on a Web site discussing treatment of ADHD suggests a strong possible bias.

Currency The site should say when it was most recently updated, and the update should be reasonable for the type of information being provided. You may need to hunt for the date; the beginning and end of the page are frequently places where updates are noted. If the document is not dated but contains a bibliography, examine the references for currency. According to the APA manual, n.d. may be used in the place of a date for a source that is not dated.

Coverage The site should thoroughly cover its topic within the limits the authors set for themselves. Coverage can be narrow or broad; the important thing is that the site does not claim to provide more information than it does or to be authoritative when it is not.

Please note that if you quote an Internet source that does not have page numbers, you need to identify the section and paragraph in which the quote appears when you cite the source in your paper.

IDENTIFY AND SUMMARIZE KEY INFORMATION FROM ARTICLES

It is not enough to find good articles; at some point you have to read them! However, reading and understanding scholarly articles is not easy. We encourage students to begin summarizing and abstracting information from articles early in the process of their literature search. Trust us, it will be easier as you go.

There are many different ways to summarize articles, and highlighting parts of the printed article is one way to start. We encourage our students to begin writing summaries of the empirical studies using the templates referred to as an **article summary sheet**, shown in Exhibit 14.2. The article summary sheet for primary, research-based studies asks you to categorize the type of study, identify the research purpose, describe the methods used in the study, summarize the

EXHIBIT 14.2
Article Summary Sheet for Primary, Research-Based Sources

a) *Type of Source*

______Journal article ______ Paper presented at conference ______Book or book chapter
______Organization report ______Government document ______Stand-alone Web site
______Other (please describe):

b) *Article Information for Reference Section*

Author(s):
(List authors in same order as in article. You only need initials of first names.)
Date of Publication*:
(*Year for journal articles and books, month and date for conference papers, newspaper, and magazine articles)
Title of Article:
Other Reference: Information**:
**Journal name: Include volume number, issue number, and page numbers. Book: Include title, publisher and city, state, and, if chapter of book, page numbers. Conference papers: Include title and city, state where conference was held. Organization reports: Include report number if available and name of organization and city, state, where located. If any of the above were obtained as full-text articles,

findings and interpretations, and list comments or criticisms that you have or quotes that you might use in your proposal. The summary does not need to be lengthy, although studies will vary in their complexity. We have not developed an article summary sheet for secondary sources and non-research-based sources because these sources vary in the information that they contain. Organize the information from these sources in a way that best suits your proposal. By writing a summary in your own words, you are beginning the process of abstracting information for your review of literature (discussed in chapter 15).

As you read and analyze the studies, you can begin to think about what other information you might need. For example, you might find out that cooperative learning has worked well with fourth graders, but results with students in earlier grades vary, with some groups doing well and others having problems. If the younger students who did well seemed to have better social skills, you might want to look for information on how training students in social skills prepares them

include name of database and date retrieved. Web site articles obtained through general search engines: include URL and date retrieved.

c) *Type of Study*

Quantitative:
_____Group experimental
_____Single subject
_____Meta-analysis
_____Nonexperimental (causal-comparative or correlational)
_____Descriptive/Survey
_____Other or I don't know (please describe)
Qualitative:
_____Ethnographic or case study
_____Narrative inquiry
_____Phenomenological
_____Other or I don't know (please describe)
Mixed-Method Research:
_____Action research
_____Mixed method
_____Program evaluation

d) *Article Summary and Critique*

Include information on participants, methods of data collection, key findings, quotes—with page numbers—and any criticisms of methods used.

for cooperative learning. As you analyze studies, you will refine your research question in ways that take into account what other researchers have found. Your revised research question might suggest new search key words to try.

WHEN DO I HAVE ENOUGH?

In every educational research class we have taught, our students want to know how many articles is enough for a research proposal. There is no magic number that automatically translates into an A grade, in part because different topics have different amounts of research. We have watched several students struggle to find more than eight sound empirical studies on looping (a teacher continues with same students for multiple years) or multiage classrooms; others find that a search on cooperative learning can turn up several hundred empirical studies. If you are finding hundreds of articles, try to refine your research question so that you can limit your search by including more combinations of variables, group characteristics, or types of settings. If you find only a few related studies, try using different keywords. Research on inclusion classrooms might yield different results if you use *inclusion* versus *mainstreaming* as your keyword. ERIC uses both *mainstreaming* and *inclusive classroom* as keywords but not *inclusion*. If the research is still limited, think about how your topic might be related to a broader one. One of our students who was studying looping found limited information in her initial searches. However, she learned that teacher-student relationships were an important issue in looping, and she expanded her literature search to include studies examining how these relationships were built over time in both looped and nonlooped classrooms as part of her literature review.

If you still want to know how many is enough, our general rule of thumb is to keep reading until you feel that the studies are saying the same things. When you feel that you are not learning anything new or that the only new studies are those cited in earlier ones you have already read, you have read enough. This might include 12 studies or 50! The expectations of what is enough may vary depending on whether you are searching literature for a master's thesis or for a paper that is one of several requirements for your course. For full-time researchers, the process of reading research continues even after data have been collected as we think about how our study fits into the vast collection of educational research that exists today. This is a good thing, because you always have something new to explore.

SUMMARY

The process of finding and researching a topic is a long journey that requires many decisions and skills. One must first generate a research topic based on experience, theory, or replication. An initial search of literature yields information on the variables and operational definitions that quantitative research on the topic has examined as well as the themes, processes, and meanings discussed in qualitative studies. Based on this review, the researcher formulates a research question that is further focused and revised as the search of literature continues.

Finding sources for a comprehensive review of literature involves multiple strategies and a wide variety of sources. Literature reviews, theoretical articles, and primary, empirical sources published in peer-reviewed journals are considered the best sources because these provide the most detailed information about research methods. However, secondary sources (for example, books, monographs), opinion pieces, and experiential reports may also be used in identifying issues or practices that warrant further study.

Strategies for finding sources include locating a review of literature; database searching using ERIC, PsycINFO, or Education Full Text; following citations to locate recent studies citing an earlier study; and using people as sources. Internet Web sites for organizations, governmental agencies, and universities and colleges may be good sources of research information. However, Internet sources should be evaluated for authority, accuracy, objectivity, currency, and coverage. A variety of engines are available for searching for information on the Internet, including some search engines that focus on academic materials.

Information from sources should be summarized with a focus on the methods, results, and conclusions reported in the studies. Article summary sheets provide a useful format for abstracting key pieces of information from articles.

The process of researching a topic may feel endless to beginning graduate students and researchers alike. Although there is no definitive answer to the question, "When do I have enough?" a rough guide is to continue researching until you feel that further studies are not yielding any new information on the variables, themes, processes, or meanings related to your topic. Professional researchers view reading research as an ongoing adventure; once a student has reviewed the major studies on his or her topic, it is time to begin developing his or her own ideas for a research study.

KEY CONCEPTS

- abstract
- article summary sheet
- citation information
- empirical research
- experiential reports
- full text
- html file
- how-to reports
- listserv
- opinion pieces
- .pdf
- peer reviewed
- personal communications
- primary sources
- replication
- researchable question
- review of literature
- secondary source
- theoretical articles

DISCUSSION QUESTIONS OR ACTIVITIES

1. Identify a topic that could be the focus of a research study, and write two or three research questions that could be addressed on this topic. Discuss whether quantitative or qualitative methods would provide the best evidence to answer each research question.
2. For the topic you select for your research proposal, locate at least two of the following types of articles: review of literature, opinion article, how-to or experiential report, empirical research study. Compare the articles you found with those found by several other students in the class. Discuss the features that can be used to identify each type of article.
3. Describe the sources and strategies that you have found most useful in researching your topic. Discuss how you revised your original research question based on what you have learned, and identify any further information that you feel is needed to complete your review of literature.

SUGGESTED READING

Pan, M. L. (2004). *Preparing literature reviews* (2nd ed.). Glendale, CA: Pyrczak Publishing.

CHAPTER FIFTEEN

THE RESEARCH PROPOSAL

CHAPTER OBJECTIVES

- Explain the difference between a research proposal and a research study
- Describe some of the key reasons for writing a research proposal
- Begin to explain in general terms the purpose and structure of the major sections of a research proposal
- Outline the key differences between a proposal written for a quantitative study and one written for a qualitative one
- Begin to conceptualize the organization of a literature review on your topic

PREPARING A RESEARCH PROPOSAL

This semester you will likely be involved in reading published research articles and writing a research proposal of your own. The research proposal in this course will be your entry into the exciting world of educational research.

A frequent source of confusion for students is the similarities and differences of published research articles and research proposals. Proposals and articles share certain characteristics but also have significant differences. Proposals and published articles follow the same general research process. That is, they identify a topic or a research question, review the literature, and develop a hypothesis or a set of subquestions. The essential difference between a proposal and a published research article is that the article is written after the study has been carried out, and it includes a "results and discussion" section in which the results of the completed study are presented and analyzed. In a published article, the researcher has carried out the study and reports findings and conclusions.

A research proposal is exactly what the term says: it is a *proposal* to conduct research. Writing a research proposal is the first step in conducting and publishing research. At any school or university, a student, staff, or faculty member conducting research must write a proposal describing his or her plans before the study can be conducted. The research proposal describes the study that you would like to conduct. It is written before the study is conducted and does not involve the collection of data. Research proposals usually have far more extensive literature reviews and method sections than do published studies.

Preparing a research proposal may seem like a daunting task at this point; this chapter is designed to help you begin to build the skills you need to develop a proposal that extends our knowledge of educational practices and policies. Many students who begin this course with little or no research experience obtain high levels of competence by the time they finish. In fact, all the research proposals included at the end of this book were written by our students. This chapter provides a broad overview of the proposal-writing process.

The overall process for writing a research proposal is similar to the process for writing an article. The skills used in writing a proposal easily translate to writing for publication. The *Publication Manual of the American Psychological Association* (APA manual) specifies the format for proposals and published studies. Our first piece of advice is to have close at hand a copy of the APA manual or a handout on APA style that may have been developed by your instructor. You will want to use the correct citation format, references format, sequence of topics, and margins. If you have looked at the APA manual, you know that we cannot review all of the specifics in the body of this book. However, in the following section of the chapter, we point out some of the key elements of a research proposal according to APA style.

Why Write a Research Proposal? There are several reasons why researchers, whether students in a research course or professional researchers, create proposals before conducting studies. Proposals describe the process and procedures that will be used by the researcher and allow an opportunity to obtain feedback from colleagues before the study is implemented. This feedback can improve the proposed study. The sharing of research designs and proposals leads to collaboration among colleagues and in many ways is a form of professional development. Each colleague brings to the discussion unique expertise, experience, and knowledge that enhance the inquiry process. Proposals are submitted to funding agencies to obtain grant money to pay for the costs of research. Master's and doctoral students may be required to obtain approval of their proposals from a committee of supervising professors. Proposals may be required from any researcher who is seeking approval from an institutional review board (IRB) committee that is reviewing the ethical issues of the study or who is submitting an article for publication. Figure 15.1 shows how the proposal fits into the process of conducting research.

In addition to providing an introduction to the topic, reviewing the literature and methods, and discussing benefits and limitations (which are described in detail following), a proposal allows the researcher to identify and address any ethical issues that are raised by the proposed study.

Anatomy of a Research Proposal

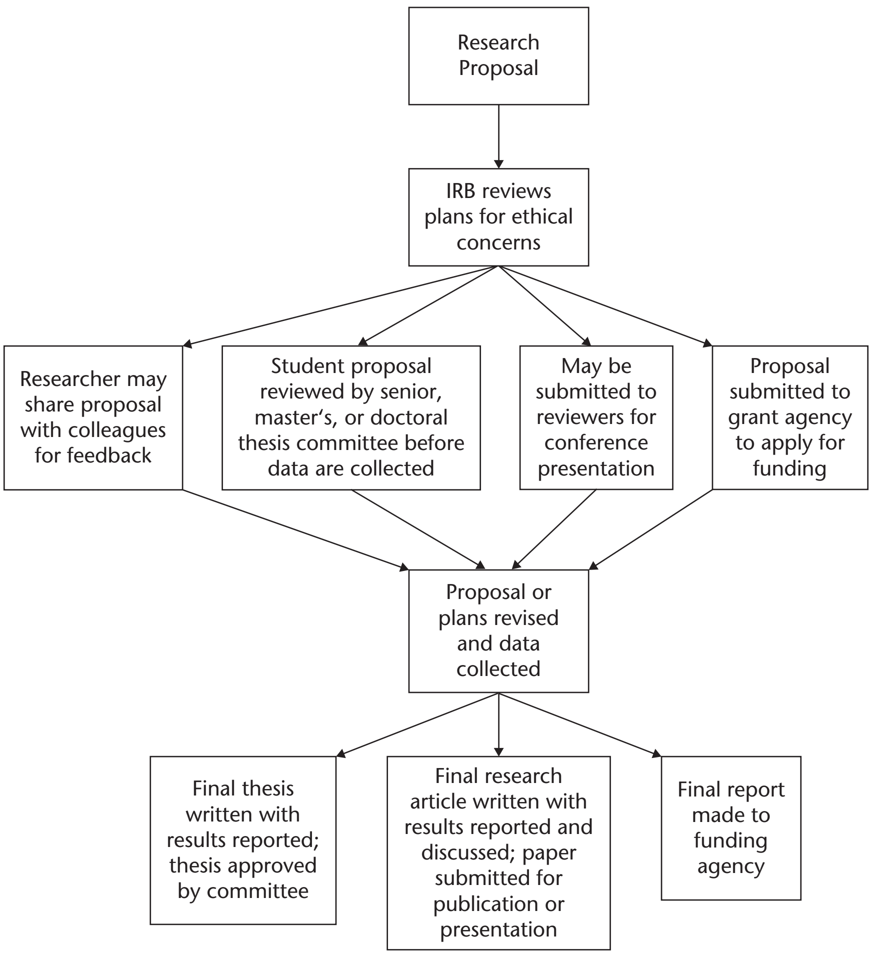
A research proposal has two overall goals. The first goal is to review past research on the topic, thereby establishing a need for the study. The second goal is to describe how the proposed study will be conducted.

The parts of the research proposal are listed here and described in the next sections.

- Title page
- Introduction
    - Statement of purpose
- Review of the literature
    - Statement of the hypothesis or subquestions
- Method

    - Participants
    - Instruments or methods of data collection
    - Design
    - Procedures
- Benefits and limitations
- References
- Appendices

FIGURE 15.1 How Research Proposals Aid the Development of Research



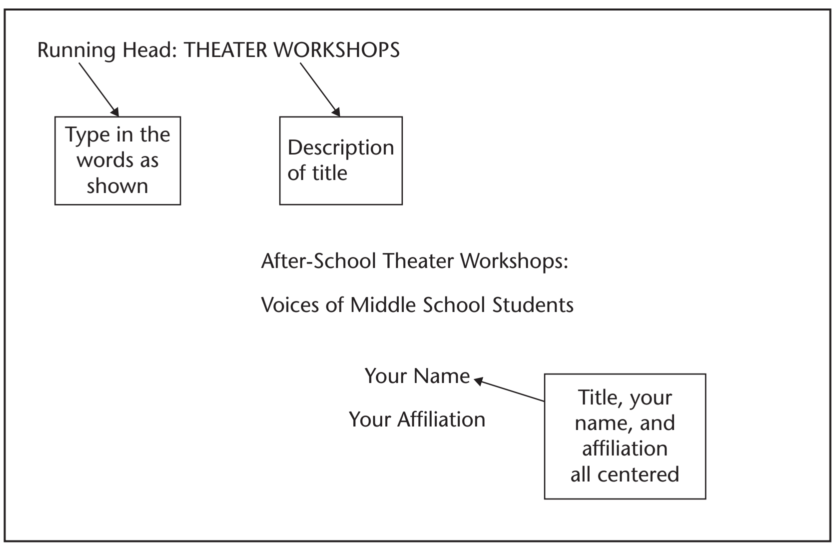
Note: IRB = institutional review board.

Title Page The **title page** is page 1 of your proposal. Several components of the title page require close attention. They include:

- *The running head:* Used in publication (but required in a proposal), the **running head** is an abbreviation of the title. It should not exceed 50 characters in length. The words "Running Head" (typed as shown in Exhibit 15.1) are flush against the left margin, followed by a colon. Two or three words selected from the title for the running head are typed in capital letters and follow the words "Running Head." On all remaining pages of the proposal only the two or three word abbreviated title appears at the top left side of each page. Page numbers should be inserted in the upper right corner of all pages.
- *The title of the proposal:* Summarizes the main focus or idea of the proposal as simply as possible. It should be clear and concise and should identify the major variables or theoretical issues to be considered in the study.
- *Author's name:* Your name and that of any coauthor are included here. If coauthors are included, you may choose to alphabetize by last name or prioritize the names in terms of the authors' contributions.
- *Affiliation:* The name of your institution (if any) is included here.

Exhibit 15.1 presents a format for a title page.

EXHIBIT 15.1 Writing Tips: Title Page



Introduction The introduction begins on page 2 of the proposal. The purpose of the **introduction** in a proposal is to provide the reader with a broad perspective of the literature and to establish a need or purpose for the study being proposed. Although it can vary in length, in essence, the introduction is a mini-version of the much larger review of literature and highlights certain segments or pieces of it (see Exhibit 15.2). As such, some researchers prefer to write the introduction after they have completed much of their literature review. In the event that you write the introduction early in the research process, you should revisit it after you have completed your literature review to be certain that there is consistency between the sections. Specifically, introductions include:

- Background information
- Definitions of variables and terms
- Statistics or contextual information that may apply to the topic
- Brief integrated summary of the findings of past research
- A brief rationale for why the proposed research study is needed

Even though the introduction to the proposal is considerably shorter than the review of literature, it too, must have citations supporting the researcher's statements concerning what is being studied. This does not mean that a citation used in the introduction cannot be used again in the review of the literature and vice versa. Citations that are mentioned briefly in the introduction can be (and usually are) discussed in detail in the complete review of literature.

Statement of Purpose The **statement of purpose** follows the introduction and describes for the reader the main purpose of the study. In an unpublished

EXHIBIT 15.2
Writing Tips: Introduction

Page 2 begins with the title of the proposal centered in the middle of the page. Body of the introduction follows. The introduction puts the study in context, provides background information, and includes a definition of variables or terms. It is basically a mini-version of your review of literature. While you may write a draft of the introduction early in the writing process, you will likely have to go back and modify it to reflect the specifics of your study.

Page 2 looks like this:

THEATER WORKSHOPS 2

Title (centered at the top of the page)

Begin writing your introduction!

research proposal, this section is usually labeled "Statement of Purpose" (see Exhibit 15.3); published studies often do not have such a definitive heading highlighting the purpose statement. Whether or not the statement of purpose is clearly labeled, it can be easily identified from its wording: "The purpose of this study was to. . . ." Although there are no universal templates for writing purpose statements, the statements should be written as clearly and as concisely as possible.

Qualitative and quantitative purpose statements are written differently. Quantitative studies clearly describe the variables that will be investigated, whereas studies that are qualitative acknowledge the emerging nature of the questions and topics. Therefore, for qualitative studies, the researcher explains that as the study progresses, the topics, variables, and questions could be changed or modified.

Review of the Literature This section of the proposal focuses on describing, summarizing, and critiquing issues, results, and explanations in the **literature** on the topic under investigation. For the purposes of this textbook, the word *literature* is defined as all published articles, records, and documents related to the topic of the study as summarized in your article summary sheets. This includes empirical research, reviews of literature, theoretical articles, and even opinion pieces (although a good literature review for a topic with ample empirical research keeps opinion pieces to a minimum). The main purpose of the literature review is to report the state of knowledge on your topic and place your proposed study in a research context by providing an overview of past research studies, published articles, and documents that relate to your topic. In writing both your introduction and your review of literature, you are becoming a member of a professional community and need to give credit to others for the work they have done. This means that you need to cite your sources whenever you include information or

EXHIBIT 15.3
Writing Tips: Statement of Purpose

You will write the phrase "Statement of Purpose" and center it in the middle of the page. A typical statement of purpose uses the following language: The purpose of the proposed study is to (describe the study's purpose here). In addition to describing the purpose of the study, include your initial rationale for doing the study here. Your rationale will be further developed at the end of your review of literature. A typical statement of purpose looks like this:

Statement of Purpose

The purpose of the proposed quantitative study is to describe teachers' attitudes about the effectiveness of inclusion classrooms. Given the popularity of inclusion classrooms and the importance of providing successful learning environments to all students, it is critical to more fully understand these issues from the teachers' perspectives.

ideas from them in your paper. Once again you should refer to the *Manual of the American Psychological Association* for citation formats.

The literature reviews in proposals and in published studies are held to different standards. You may have noticed that the literature reviews of published studies are often brief, condensed into a few paragraphs or pages, and are sometimes included as part of the introduction; however, in a research proposal, the review of literature is exhaustive and is separate from the introduction.

From this review of previous research, one is pushed to reflect on one's research question and often ask new or modified questions. Reviews of literature are well planned out, typically starting at the broadest part of the topic and gradually refining and narrowing the discussion down to a more focused perspective (Exhibit 15.4).

Statement of the Hypothesis or Subquestions The literature review ends with either a research hypothesis and operational definitions of variables or a refined research question and a set of subquestions.

Research Hypothesis and Operational Definitions For experimental, causal-comparative, and correlational studies, this section of the proposal includes the research hypothesis or prediction of the expected outcome of the study. It is by no means an accident that the hypothesis comes directly after the review of literature. Hypotheses are based on and consistent with the findings reported in the literature review. Hypotheses are always accompanied by operational definitions of variables because these indicate how the variables will be measured or manipulated. Operational definitions may be based on methods used by the studies in the review of literature. However, it is important that the researcher include them at this point to make it clear how the variables will be defined in the study being proposed. If done well, a review of literature leads the reader to make the same hypothesis—or, at a minimum, to see the logic in your hypothesis.

Research hypotheses for experimental research studies predict how the independent variable(s) will affect the dependent variable(s) in the proposed study. All levels of the independent variable need to be defined.

Research Question and Subquestions In all types of qualitative research and in descriptive survey research, a research question and a set of subquestions rather than a hypothesis follow the review of literature. Although descriptive survey research is among the family of quantitative research methods, it tends to be more exploratory and usually does not include a hypothesis. Instead the researcher writes a set of subquestions that form the basis for the survey.

EXHIBIT 15.4
Writing Tips: The Review of Literature

Immediately following the statement of purpose, you will begin your literature review. Type the words “Review of the Literature,” and center the phrase. Do not start the literature review on a separate page. It should look like this:

Review of Literature

The literature review begins with a brief introduction (see example below). Here you may want to introduce the literature review by indicating the strands of literature you will be covering and the type of research you found on your topic.

> **Example introductory portion of literature review:**
> This paper provides a review of the literature on teachers’ attitudes toward inclusion and describes many important issues relative to the success of this practice for special and regular education students. Much of the research in this area has been quantitative, using a descriptive survey approach.

Although you will not provide detail on all of the articles you have reviewed, you will provide information related to the study findings and methods. If you have followed our suggestions, you have summary sheets (or the equivalent article abstracts) that summarize the articles related to your topic. Remember, literature reviews are organized by themes and issues and are challenging to write. Many students want to simply review each article as though they were writing a cookbook listing one ingredient after another. This is exactly what you do not want to do! What follows are the steps you should consider when writing your literature reviews:

Step 1. Examine your article summary sheets or article abstracts and highlight or underline the purpose of the study, the methods used, and the findings. Then, as you read over your summary, articulate a summary word, topic, or phrase that best applies to the article. Write this on the top of the article summary sheet. For some articles, there may be several summary topics or phrases. You will use these topics or phrases later as means to organize your review.

Step 2. Review the summary words you have generated. It is likely that because the research is all related, there will be overlap in summary words you have generated. Now place each article under the summary word or phrase that best fits. You should now have several piles of articles categorized by their essential focus. Some articles will likely fit in more than one category.

Step 3. Decide how your literature review should be organized. Use the following questions to help guide your organization:

- How can I create a complete picture of what is known about my research focus?
- What studies are most related to my proposed study and least related?
- Which studies support my hypothesis or subquestions, and which studies conflict with it?
- How can I create a meaningful framework for my study (lead the reader to the conclusion that my study will answer an unanswered question or provide valuable information to the field, or both)?

Step 4. Begin to write your literature review using the organization you created for your topics. Following your introductory comments, you might want to begin with something like the following example:

> **Example transition from introductory comments to body of the literature review:**
> One issue that has emerged with frequency in the literature on inclusion is the degree to which teachers believe they have been adequately prepared to work with both special and regular education students. For example, Tomas (2004) found that many regular education teachers believed that they should have taken more courses that deal with special ability children.

(*Continued*)

EXHIBIT 15.4 (*Continued*)

Now include in this section of the literature review a discussion of all of the articles that have investigated the same topic (in this example teacher preparation for inclusion). Continuing with the example, you might transition to the next article by saying something like:

> Similar findings were reported by Johnson (2003). Johnson's study with 25 regular education teachers found that most believed that their special education training was not adequate.

Once you have exhausted the articles on that topic, you can then move on to any other topics you have identified in your articles. Be sure to make clear transitions between the sections as noted in bold below. For example:

> **In addition to teachers' attitudes toward their preparedness for inclusion**, researchers who are interested in teachers' attitudes toward inclusion have addressed teachers' perception of the impact of inclusion on the achievement of special and regular education students once they are placed in an inclusion classroom.

Continue this process until you have included all of the articles you believe are critical to your topic. You then conclude the literature review with a brief summary statement that makes connections among the various studies on the topic. Focus on pointing out consistencies or conflicts in methods used or results obtained (or both) among the studies you described. You want to say something like:

> The research in the area of teacher attitudes toward inclusion has demonstrated that many teachers have concerns about the inclusion classroom. Specifically, they are concerned with . . .

Qualitative approaches use research questions and subquestions that guide the qualitative researcher during the initial investigative process. The research question describes the overall focus of the study, which is then broken down into subquestions that are addressed in the study. (They are called *subquestions* because they break down the broad research question and guide the collection of data without predetermining outcomes.) Both the research question and the subquestions should be connected with the purpose statement and suggest the types of procedures that are further explored in the method section. Remember, subquestions are not your actual interview questions, but they do guide interviews or observations.

Exhibit 15.5 presents information about preparing the hypotheses or subquestion part of the research proposal.

EXHIBIT 15.5
Writing Tips: Hypotheses or Subquestions

EXPERIMENTAL RESEARCH HYPOTHESES

If your proposal includes a hypothesis, it is critical that the hypothesis be written clearly and concisely and be based on the literature review. The types of hypotheses are covered in chapter 9, but all hypotheses include the following information:

- The predicted outcome of the study, including all variables being measured and manipulated
- A clear identification of all groups in the study
- An operational definition of the dependent variable

An example of a hypothesis written for an experimental design follows. If you decide to do an experimental study, your study should have a similar format. Causal-comparative research also follows this format.

Sample Research Hypothesis for an Experimental Research Study

It is hypothesized that the science achievement scores of 10th graders will be higher for groups using portfolios than for those using separate assignments. For the purpose of this proposed study, the dependent variable is science achievement and is operationally defined as students' scores on an end-of-the-year 10th-grade science curriculum assessment.

QUALITATIVE RESEARCH SUBQUESTIONS

In qualitative research, the initial research question is typically broken down into subquestions that can be used to guide observations or interviews. However, the researcher needs to be careful not to limit what the study may find. Questions may be modified throughout the course of the study.

If your proposal includes subquestions, you must be certain that the questions are specific and will serve as a guide for the method. They should include:

- The specific questions that will be addressed in the study
- The way in which the data will be collected

An example of subquestions written for a case study or ethnography follows.

Sample Research Subquestions for a Case Study or Ethnography

The initial research question to be explored in this study is, *What changes does this student have to make when the focus of his or her instruction changes from*

(*Continued*)

EXHIBIT 15.5 (*Continued*)

a lecture-based teacher-centered approach to a more student-centered alternative assessment method of instruction?

Possible subquestions to be addressed in this study include:

1. What decisions does the student have to make regarding his approach to studying the course material?
2. How does the student plan for what materials should be included in the alternative assessment (e.g., portfolios)?
3. How does the more student-centered alternative assessment change the student's relationship with the teacher?

Method The **method section** focuses on the study proposed and describes how the researcher plans to carry out the study. Because it describes what you will do, this section is written in the future tense. It is broken into subsections as follows. Note that the subsections differ depending upon whether one is proposing a qualitative or quantitative study. Action research studies typically follow the format for qualitative studies, since they usually include at least one qualitative measure.

Participants A description of the **participants**, often referred to as the *sample*, is the first subsection found in the method section of a proposal. Methods used to select the participants (otherwise known as *sampling techniques*) are discussed here, as are the characteristics of the proposed sample. Keep in mind that for many studies, a more in-depth description of the sample (for example, age, years teaching, and so on) is developed during the data collection efforts through the gathering of demographic information (for example, by a survey). In such cases, the participant subsection can be written up with more detail and inserted later, after the study has been completed. However, in the proposal you should describe the type of demographic data that you plan to collect and how you will obtain it. In some cases, participant demographic data can be collected from government agencies, school report cards, or associations that collect and store this kind of information. Many of these organizations provide access to these data through Web pages.

Researchers who use quantitative approaches may find that such demographic data is more available (and applicable) than those researchers who are using qualitative approaches. Qualitative demographic data is more context specific and often requires more "undercover" work by the researcher. In qualitative proposals, the participant section describes why this setting or these participants qualify as key informants. *Key informants* are individuals who have some unique information or knowledge of the phenomena being studied. These descriptions often provide rich detail about the setting and participants and how the key informants will help to answer the research questions. Both quantitative and qualitative proposals must indicate how many participants you expect to have in your study.

Instruments or Methods of Data Collection For this subsection, quantitative researchers use the term **instruments**, and qualitative and action researchers use the phrase **methods of data collection**. The approach you select for your proposal (for example, quantitative or qualitative) plays a key role in the instruments you will propose to use. Traditionally, quantitative research relies more on standardized instruments to collect data, and qualitative approaches use what are called *self-developed interview* or *observation* (or both) *protocols*. Regardless of the approach, this subsection does not describe how the researcher will use the instrument but rather indicates what the instrument or protocol "looks like" and the reasons or support for selecting it—in other words, how it will generate the data needed to answer the research question. Those researchers who are conducting quantitative research using standardized, already established instruments must include in this subsection critical reliability and validity evidence that supports the use of these instruments. (See chapter 4 for more in-depth information on reliability and validity in measurement.)

For researchers creating their own instruments (for example, a survey or interview protocol) or conducting qualitative research, this subsection is called the "Methods of Data Collections" and includes the researcher's description of the surveys, observational checklists, or protocols that will guide observations or interviews. For a survey, a description including how many sections the survey has and what type of items are used is included along with the survey questions in an appendix to the proposal. For interview or observation protocols, a similar narrative defining the purpose and the subquestions guiding the observations or interviews are included in the text, and the protocols are included in appendices.

Design The **design** section of the proposal is traditionally included only for experimental research. In this subsection, the researcher discusses the model or design that he or she has chosen and how threats to the study's validity are controlled by the design (these topics are addressed in detail in chapter 9 on experimental research). Usually, a short narrative accompanied by a table or figure that visually supports the narrative is included. The table or figure alone should provide a clear sense of the design the researcher intends to use (see Table 15.1 for an example of an experimental design).

TABLE 15.1 Sample Design

Group	Pretest Data	Treatment	Posttest Data
A	Iowa math scores from prior year	A (use math software program)	Iowa math scores at end of year
B	Iowa math scores from prior year	B (individual tutoring without computers)	Iowa math scores at end of year

Procedures In proposals for which a nonexperimental research design is planned, the procedure subsection immediately follows the instruments subsection, and the type of research proposed (for example, correlational, case study) is identified at the start of the procedure subsection.

The **procedures** subsection is often referred to as the "how-to" section. A narrative describes for the reader the process and practices that the researcher plans to use to collect the data, using methods that are appropriate for answering the research questions. In this section, the researcher describes how the instrument(s) or data collection tools will be used. Qualitative researchers generally discuss how they anticipate approaching the selected research setting and the tactics or strategies they anticipate using as they ask initial questions and gather preliminary data. Most important, the role of the researcher and his or her relationship to the participants are clearly outlined. Quantitative researchers discuss the timeline of the project, when instruments will be given to participants to collect data, what they are told about the study, and who conducts each phase of the study. Quantitative researchers also provide a detailed description of the treatments and how they are delivered.

Keep in mind that not everything always works out as planned, particularly in qualitative studies. Researchers often go back after they have conducted the study and modify their procedure subsection to accurately reflect the process that occurred. Note that although modification in the procedures is often expected in qualitative research, such modifications and deviations from the initial procedures are limited in quantitative research.

See Exhibit 15.6 for detailed instructions on preparing the method section of a research proposal.

Benefits and Limitations In this section of the proposal, the researcher discusses both the **benefits** and the **limitations** of the research that he or she is proposing. Benefits typically are twofold: one aimed at the benefits of the research study for advancing knowledge represented by the research literature and the other targeting the benefits for practice. The researcher wants to describe how the study

EXHIBIT 15.6
Writing Tips: Method Section

The method section starts on a separate page, and the heading "Method" must be typed and centered as shown below.

Method

The method section is divided into several subsections that are labeled, italicized, and placed flush against the left margin. Note that the **proposal should be in paragraph form.** Although we use bullets to identify the type of information included in each subsection, bullets are not used in the proposal. Remember to use the future tense. Do not begin a separate page for each of the subsections.

Participants
Include the following in the participants section:

- Description of the demographic background of the participants (this could include race, ethnicity, age, and gender or other information relevant to the study, such as reading level or classification of disabilities)
- Description of how the participants will be selected for the study
- The number of participants to be included (may change later for qualitative research)
- Discussion of ethical issues and how informed consent will be obtained

Methods of Data Collection or Instruments
When writing your methods of data collection section, describe in detail how the data will be collected and what tools and instruments will be used to collect the data. Specifically, for quantitative studies you should

- Describe the measuring instruments (the survey, test, or behavior checklist or the type of archival data to be collected)
- Provide a rationale for using these instruments (a discussion of how the data obtained from these instruments will provide information to answer the research question)
- Describe how the instruments measure the variables
- Report the reliability and validity of the instruments
- For researcher-developed instruments, describe the construction and validation of the instrument and provide a copy of the instrument in an appendix

If your study is qualitative, this section should

- Describe in detail the data collection tools (observations or interview protocols or the type of archival data to be collected)
- Describe completely how the data will be collected and by whom
- Describe any triangulation methods you will use to establish credibility of the procedures you have proposed
- Describe methods used to establish dependability of your methods of data collection
- Discuss development of the interview or observation protocol and provide a copy in the appendix

Materials
Not all proposals include a materials section. This section is specifically for studies that include curriculum materials, instructional materials, or any materials, other than measuring tools, that the researcher plans to use as part of the method. Often materials include things like books, workshop materials, computers, instructional software, and the like

Design
When writing the design section for experimental research, be sure to describe the kind of study you are proposing. Specifically, you should

- Describe the approach of the study. Is it quantitative? If so, what kind? Is it qualitative? If so, what kind? Is it a mixed-method or action research study?
- Include a description of the number of groups in the study and whether they are treated differently

Procedure
A good procedure section includes the following:

- A description of the role of the researcher (detached observer, interviewer, teacher, and so on)
- Detailed description of the procedures you propose for your study. This description must be sufficiently detailed to allow another person to conduct and replicate the proposed study
- Use the researcher's point of view to describe how the study will be organized and the participant's point of view to describe the task

will help support and extend understanding of the topic. Typically, this involves examination of how your research will add to or answer unanswered questions from the last few studies in the literature review (with citations), because these, by their very nature, should be the most closely related pieces of work.

This section includes a short paragraph summarizing some of the possible implications of the hypothesized findings for potential benefits to teaching and learning. The final paragraphs of this section focus on limitations of the proposed research. Such limitations tend to focus on methodological limitations. A professional researcher is always aware of the limitations of the methods used in a study and any possible criticisms that might be made of the study by other researchers or practitioners. While a good proposal includes proposed methods that are as strong as possible, there is no perfect study. This section provides an opportunity to reflect on what you have proposed and offer some final self-criticisms regarding problems that remain.

References The **references** section contains all the publication information for citations used in the proposal. Unlike a bibliography, which lists all references encountered during the research process (including some that may not have been cited), the reference page includes only those references that were cited somewhere in the paper. Most of the citations from the references appear in the review of literature, but citations may be used to support ideas or define variables at any part of the proposal. Reference sections of proposals must be in APA style. For further clarification of the APA style, refer to the most recent edition of the *Publication of the American Psychological Association*.

Appendices All researcher-developed tools are included in the **appendix** section of the proposal (for example, the informed consent letters, the final form of any survey, interview protocol, observational protocol, or materials). If multiple instruments are developed, each instrument is included in a separate appendix, and each appendix is given a letter label (for example, Appendix A, Appendix B). These labels are used when referencing the item in the body of the proposal. Note that all appendices should be referred to in the text.

Samples of quantitative and qualitative research proposals written by our students are included in Appendixes A and B of this book.

SUMMARY

Although research proposals and published articles have similar characteristics, the major difference is that a proposal describes a plan for a study that has not yet been carried out. A research proposal is necessary for several reasons: for an ethical review of the proposed research and methods by an institutional review board, to elicit feedback and critique from colleagues and peers, and to provide a framework for the study when presenting it to possible participants or administrators of schools or settings where the research will take place. To provide continuity across studies, researchers adhere to the guidelines presented by the *Publication Manual of the American Psychological Association* (APA). Research proposals should be written in APA style.

The research proposal is composed of several sections. The introduction is the first section. It contains the statement of purpose, and the study's rationale. This is followed by the review of literature and the statement of the hypothesis or subquestions. Next is the method section that contains the participants subsection, instruments or methods of data collection, design, procedures, and finally the benefits and limitations section. References are at the end of the proposal, starting on a separate page. If the researcher has developed an informed consent letter or instruments such as a survey or interview protocol, such material is included in an appendix following the reference section.

KEY CONCEPTS

appendices
benefits and limitations
design
header
instruments
introduction
literature
method section
methods of data collection
participants
procedures
references
running head
statement of purpose
title page

DISCUSSION QUESTIONS OR ACTIVITIES

1. You are writing a research proposal on the charter school movement or other topic of your choice. Discuss in what ways your proposal would differ if you wrote it from a quantitative and then a qualitative perspective.
2. Discuss why qualitative researchers end their reviews of literature with subquestions whereas quantitative researchers generate research hypotheses.
3. Some researchers, especially students in an educational research course, might argue that a research proposal is an unnecessary step in the research process. Discuss in specific terms why a proposal would be important to the researcher, the institution where the research is being conducted, and to the participants in the following study: What is the effect of peer tutoring on the math achievement of at-risk third grade students?

SUGGESTED READINGS

Davitz, J. R., & Leiderman Davitz, L. (1996). *Evaluating research proposals: A guide for the behavioral sciences.* Upper Saddle River, NJ: Prentice Hall.

Pyrczak, F., & Bruce, R. R. (1992). *Writing empirical research reports: A basic guide for students of the social and behavioral sciences.* Los Angeles: Pyrczak Publishing.

REFERENCES

Allport, G. (1955). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher, 27*(2), 65–75.

Andersen, R. E., Crespo, C. J., Bartlett, S. J., Cheskin, L. J., & Pratt, M. (1998). Relationship of physical activity and television watching with body weight and level of fatness among children. *Journal of the American Medical Association, 279*(12), 938–943.

Anderson, D. D. (2008). The elementary persuasive letter: Two cases of situated competence, strategy, and agency. *Research in the Teaching of English, 42*(3), 270–314.

Anderson, G. L., Herr, K., & Nihlen, A. S. (1994). *Studying your own school: An educator's guide to qualitative practitioner research*. Thousand Oaks, CA: Corwin.

Arhar, J. M., Holly, M. L., & Kasten, W. C. (2001). *Action research for teachers: Travelling the yellow brick road*. Upper Saddle River, NJ: Prentice Hall.

Bandura, A., Grusec, J. E., & Menlove, F. L. (1966). Observational learning as a function of symbolization and incentive set. *Child Development, 37*(3), 499–506.

Baumrind, D. (1967). Child care practices anteceding three patterns of preschool behavior. *Genetic Psychology Monographs, 75*(1), 43–88.

Brookhart, S. (n.d.). Review of Iowa Test of Basic Skills. In O. K. Burros (Ed.), *Mental measurement yearbook on-line*. Retrieved May 10, 2005, from Silverplatter database: http://spweb.silverplatter.com.

Boggs, R. (n.d.). *Exploring data: Olympic gold*. Retrieved from http://Exploringdata.cqu.edu.au/datasets/oly_gold.xls.

Bracht, H. G., & Glass, V. G. (1968). The external validity of experiments. *Journal of the American Educational Research Association, 5*(4), 437–474.

Brown, K. (1998). Ladders to literacy. *Threshholds in Education, 24*(3), 4–10.

Bruchac, J., & Bruchac, J. (1998). *When the Chenoo howls: Native American tales of terror*. New York: Walker & Co.

Camic, P. M., Rhodes, J. E., & Yardley, L. (2003). Naming the stars: Integrating qualitative methods into psychological research. In P. M. Camic, J. E. Rhodes, & L. Yardley (Eds.), *Handbook* *of qualitative research in psychology*: *Expanding perspective in methodology and design* (pp. 3–16). Washington, DC: American Psychological Association.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Campbell, D. T., & Stanley, J. S. (1971). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Carger, C. L. (1996). *Of borders and dreams: A Mexican-American experience of urban education*. New York: Teachers College Press.

Case, R. (1992). Neo-Piagetian theories of child development. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 161–196). New York: Cambridge University Press.

Chalk, J. C., Hagan-Burke, S., & Burke, M. D. (2005). The effects of self-regulated strategy development on the writing process for high school students with learning disabilities. *Learning Disability Quarterly, 28*(1), 75–87.

Chambliss, D. F., & Schutt, R. K. (2003). *Making sense of the social world: Methods of investigation*. Thousand Oaks, CA: Sage.

Chase, S. E. (2005). Narrative inquiry: Multiple lenses, approaches and voices. In N. Denzin & Y. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 651–679). Thousand Oaks, CA: Sage.

Ciechalski, J. E. (n.d.). Review of Mathematics Self-Efficacy Scale. In O. K. Burros (Ed.), *Mental measurement yearbook on-line*. Retrieved from Silverplatter database: http://spweb.silverplatter.com.

Clandinin, D. J., & Connelly, F. M. (1988). *Teachers as curriculum planners: Narratives of Experience*. New York: Teachers College Press.

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture*. Berkeley: University of California Press.

Cochran-Smith, M., & Lytle, S. L. (1993*). Insider/outsider: Teacher research and knowledge*. Language and Literacy Series. New York: Teachers College Press.

Cohen, L., & Manion, L. (1994). *Research methods in education*. London: Routledge.

Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researchers, 19*(5), 2–14.

Cook, T. D., Campbell, D. T., & Peracchio, L. (1979). Quasi experimentation. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., 491–576). Palo Alto, CA: Consulting Psychologists Press.

Covin, J., & Dirkx, J. M. (2003). "Being called awake": The role of transformative learning in the lives of environmental activists. *Adult Education Quarterly, 53*(2), 99–118.

Craig, C. (2003). *Narrative inquiries of school reform: Storied lives, storied languages, storied metaphors*. Research in Curriculum and Instruction. Greenwich, CT: Information Age.

Craig, D. V. (2009). *Action research essentials*. San Francisco: Jossey-Bass.

Creswell, J. (2002). *Educational research: Planning, conducting and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson Education.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Merrill Prentice Hall.

Cronk, B. C. (2008). *How to use SPSS: A step-by-step guide to analysis and interpretation*. Glendale, CA: Pryzcrk Publishing.

Darling-Hammond, L. (2001). *The research and rhetoric on teacher certification: A response to "Teacher Certification Reconsidered."* Retrieved from ERIC database. (ED 477296)

Davies, R. D. (2003, November). *To prove or improve*. Paper presented at the American Evaluation Association, Reno, NV.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological method*. New York: Praeger.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2005). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Dillman, P., & Salant, D. A. (1994). *How to conduct your own survey*. New York: Wiley.

Doran, H. C., & Fleischman, S. (2005). Challenges of value-added assessment. *Educational Leadership, 63*(3), 85–87.

Duquette, C., Stodel, E., & Fullarton, S. (2006). Persistence in high school: Experiences of adolescents and young adults with fetal alcohol syndrome. *Journal of Intellectual and Developmental Disability, 31*(4), 219–231.

Ensler, E. (1998). *The vagina monologues: The V-day edition*. New York: Villard/Random House.

Fitzpatrick, B. L., Sanders, J. R., & Worthen, J. F. (2004). *Program evaluation: Alternative approaches and practical guidelines*. Boston: Allyn & Bacon.

Flores, K. S. (2008). *Youth participatory evaluation: Strategies for engaging young people*. San Francisco, CA: Jossey-Bass/Wiley.

Frechtling, J. A. (2007). *Logic modeling methods in program evaluation*. San Francisco, CA: Jossey-Bass/Wiley.

Freire, P. (1970). *Pedagogy of the oppressed*. (M. B. Ramos, Trans.). New York: Herder & Herder.

Friend, J. H., & Guralnik, D. (1956). *Webster's new world dictionary of the American language*. New York: World.

Gall, M. D., Gall, W. R., & Borg, J. P. (2005). *Educational research: An introduction* (6th ed.). White Plains, NY: Longman.

Garcia, J. (2009, May). *Photovoice: A participatory action research method for empowerment and social justice*. Workshop presented at the College of St. Rose, Albany, NY.

Gay, L. R., & Airasian, P. (2003). *Educational research: Competencies for analysis and application*. Columbus, OH: Prentice Hall.

Gay, L. R., Mills, G. E. , & Arasian, P. (2006). *Educational research: Competencies for analysis and application* (8th ed.). Upper Saddle River, NJ: Prentice Hall.

Geer, C., & Hamill, L. (2007). An online interdisciplinary discussion: Promoting collaboration between early childhood and special education preservice teachers. *Journal of Technology and Teacher Education, 15*(4), 533–553.

Glesne, C. (2006). *Becoming qualitative researchers: An introduction* (3rd ed.). Boston: Allyn & Bacon.

Go, C., & Murdock, S. (2003). To bully-proof or not to bully-proof: That is the question. *Journal of Extension, 41*(2). Retrieved from http://www.joe.org/joe/2003april/rbl.shtml.

Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and qualitative design in educational research*. New York: Academic Press.

Gold, R. L. (1958). Roles in sociological field observation. *Social Forces, 36,* 217–223.

Gray, D. E. (2004). *Doing research in the real world*. Thousand Oaks, CA: Sage.

Harter, S., & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*(6), 1969–1982.

Hoffman, L. M. (2004). We're so diverse: How students use their high school yearbooks to bridge the gaps. *American Secondary Education, 33*(1), 4–25.

Hubbard, R. (1989). Notes from the Underground: Unofficial literacy in one sixth grade preview. *Anthropology and Education Quarterly, 20*(4), 291–307.

Humphrey, D. C., Wechsler, M. E., & Hough, H. J. (2008). Characteristics of effective alternative teacher certification programs. *Teachers College Record, 110*(1), 1–63.

Jordan, L., & Hendricks, C. (2002). Increasing sixth grade students' engagement in literacy learning. *Networks, 5*(1). Article 1. Retrieved June 26, 2009, from http://journals.library.wisc.edu/index.php/networks/issue/view/19.

Kennedy, C. H. (2005). *Single-case designs for educational research.* Boston: Allyn & Bacon.

Kennedy, J., & Voegtle, K. (2000, Sept.). Friendship merging with ideology: The process of becoming an ally. *Initiatives* [Online serial]. Retrieved from http://nawe.org/members/journal/59.4/kennedy.htm.

Kidder, T. (1989). *Among schoolchildren.* Boston: Houghton Mifflin.

King, B. (1990, March). *Creating curriculum together: Teachers, students, and collaborative investigation.* Paper presented at the annual meeting of the American Educational Research Association, Boston. Retrieved from ERIC database. (ED 322111)

Klaw, E. (2008). Understanding urban adolescent mothers' visions of the future in terms of possible selves. *Journal of Human Behavior in the Social Environment, 18*(14), 441–461.

Kozol, J. (1995). *Amazing grace: The lives of children and the conscience of a nation.* New York: HarperCollins.

Kozol, J. (2000). *Ordinary resurrections: Children in the years of hope.* New York: Crown.

Larson, R. (1989), Beeping children and adolescents: A method for studying time and use and daily experience. *Journal of Youth and Adolescence, 18*(6), 511–530.

Lau, M. Y., Sieler, J. D., & Muyskens, P. (2006). Perspectives on the use of the problem-solving model from the viewpoint of a school psychologist, administrator, and teacher from a large midwest urban school district. *Psychology in the Schools, 43*(1), 117–127.

LeCompte, M. D., & Schensul, J. J. (1999). *Designing and conducting ethnographic research: Ethnographer's toolkit* (Vol.1). Walnut Creek, CA: AltaMira.

Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends.* Cambridge, MA: Civil Rights Project at Harvard University (The). Retrieved from http://www.civilrightsproject.ucla.edu/research/esea/nclb_naep_lee.pdf.

Lewin, K. (1948). *Resolving social conflicts; selected papers on group dynamics.* G. W. Lewin (Ed.). New York: Harper & Row.

Lincoln, Y. S. (1995). In search of students' voices. *Theory into Practice, 34*(2*),* 88–93.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.

Linn, R. L. (2003). Performance standards: Utility for different uses of assessment. *Education Policy Analysis Archives, 11*3(1). Retrieved from http://epaa.asu.edu/epaa/v11n31/

Lodico, M. G., & Voegtle, K. H. (2005). *Child and adolescent life stories: Perspectives from youth, parents, and teachers.* Thousand Oaks, CA: Sage.

Malesky, L. A. Jr., & Ennis, L. (2004). Supportive distortions: An analysis of posts on a pedophile internet message board. *Journal of Addictions and Offender Counseling, 24*(2), 92.

Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of Personality and Social Psychology, 3*(5), 551–558.

McClelland, J. (1997). Knowing and being known: Parents' experiences with rural schools. *Journal of Research in Rural Education, 13*(2), 108–116.

McNiff, J. (1988). *Action research: Principles and practice*. Basingstoke, UK: Macmillan.

McTaggart, R. (Ed.). (1997). *Participatory action research*. Albany: State University of New York Press.

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Wiley.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Mills, G. E. (2000). Action research: Accountability, responsibility and reasonable expectations. *Educational Researcher, 32*(7), 3–13.

Mills, E. G. (2003). *Action research: A guide for the teacher researcher* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Mills, L. J., & Daniluk, J. C. (2002). Her body speaks: The experience of dance therapy for women survivors of child sexual abuse. *Journal of Counseling and Development, 80*(1), 77–85.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for education reform in America*. Retrieved from http://www.ed.gov/pubs/NatAtRisk/index.html.

National Association of School Psychologists. (n.d.). *Position statement on prevention and intervention research in the school*. Retrieved from http://www.nasponline.org/about_nasp/pp_prevresearch.aspx.

Neuman, S. (2002). *Proven methods: Scientific based research*. Seminar transcripts. Retrieved from http://www.ed.gov/nclb/methods/whatworks/research/index.html.

Newman, D. L., & Spaulding, D. T. (1997). *Research in the high schools: National Science Foundation*. Albany: State University of New York, Evaluation Consortium.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high stakes testing corrupts America's schools.* Cambridge, MA: Harvard Educational Publishing Group.

Ogbu, J. U. (2003). *Black American students in an affluent suburb: A study of academic disengagement*. Mahwah, NJ: Lawrence Erlbaum.

On-Line Learning Center. (2005). *Cooperative learning lesson*. Retrieved from http://olc.spsd.sk.ca/DE/PD/instr/strats/coop/lesson.pdf.

Pasko, M. (2004). Curriculum connections: Linking literature and math. *Networks, 7*(2). Retrieved from http://education.ucsc.edu/faculty/gwells/networks/journal/Vol.7(2).2004may/Pasko.html.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text.* Retrieved from ERIC database. (ED 413355)

Perkins, T. S., Ayala, J., Fine, M., & Erwin, E. J. (n.d.). *Ethnographic study of "Playtime Is Science" for children who are blind or visually impaired: "You don't have to be sighted to be a scientist, do you?"* Retrieved from Educational Equity Center Web site at: http://www.edequity.org/ethnographicstudy.pdf.

Personal Narratives Group. (1989). *Interpreting women's lives.* Bloomington: Indiana University Press.

Phillips, D. C., & Burbules, N. C. (2000). *Positivism and educational research*. Lanham, MD: Rowman & Littlefield.

Polanyi, M. (1967). *The tacit dimension*. Chicago: University of Chicago Press.

Pritchard, I. A. (2002). Travelers and trolls: Practitioners and institutional review boards. *Educational Researcher, 31*(3), 3–13.

Proctor, M. H., Moore, L. L., & Gao, D. (2003). Television viewing and change in body fat from preschool to early adolescence: The Framingham children's study. *International Journal of Obesity and Related Metabolic Disorders, 27*(7), 827–833.

Quinn, G. P., Hauser, K., Bell-Ellison, B. A., Rodriguez, N. Y., & Frías, J. L. (2006). Promoting pre-conceptional use of folic acid to Hispanic women: A social marketing approach preview. *Maternal and Child Health Journal, 10*(5), 403–412.

Raider Open Door Academy. (n.d.). Literacy skills checklist. Retrieved from http://nettleton.crsc.k12.ar.us/~ncovey/Literacy%20Skills%20Checklist.htm.

Reason, P. & Bradbury, H. (2008). *Handbook of action research* (2nd ed.). Thousand Oaks, CA: Sage.

Runswick-Cole, K. (2007). "The tribunal was the most stressful thing: More stressful than my son's diagnosis or behavior"—The experiences of families that go to the special educational needs and disability tribunal. *Disability and Society, 22*(3), 315–328.

Salant, P., & Dillman, D. A. (1994). *How to conduct your own survey.* Hoboken, NJ: Wiley.

Sandall, N., Schramm, K., & Seibert, A. (2003*). Improving children's listening skills through the use of literature.* Masters of Arts Research Project, Saint Xavier University, Chicago, IL. Retrieved from the ERIC database (ED482002)

Schellenberg, S. J. (2004, April). *The achievement gap: Test bias or school structure?* Paper presented at the Annual Meeting of the National Council for Measurement in Education. Retrieved from http://datacenter.spss.org/sites/2259653e_ffb.../text_bias_paper.pdf.

Schmitt-Rodermund, E., & Vondracek, F. W. (1998). *Breadth of interests, exploration, and identity development in adolescence*. Retrieved from ERIC database. (ED 420 445)

Schneider, M. (2006). *Comment delivered to the American Enterprise Institute for Public Policy Research*. Retrieved from US Department of Education, Institute of Education Science Web site: http://nces.ed.gov/whatsnew/commissioner/remarks/2_6_2006.asp.

Schön, D. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.

Schwandt, T. A. (2007). *The SAGE dictionary of qualitative inquiry*. Thousand Oaks, CA: Sage.

Seashore, H. G. (1980). *Method of expressing test scores.* In *Test service notebook* (p. 148). New York: Psychological Corporation.

Shepardson, D. P., Britsch, S. J. (2000, April 28–May 1). *Young children's representations of earth materials on the science journal*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, New Orleans, LA. (ED439945)

Spaulding, D. T., & Lodico, M. G. (2004). *Proposal for a feasible study for a Nativity/San Miguel model middle school*. Unpublished manuscript. Albany, NY: The College of Saint Rose.

Spaulding, D. T., Lodico, M. G., Jones, K., & Gligora, M. (2004, April). *Is a picture really worth a thousand words? Using student generated photographs and journals to document experiences in a summer enrichment program for at-risk youth*. Paper presented at the American Educational Research Association Conference, San Diego.

Stake, R. E. (2000). In N. K. Denzin and Y. K. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 435–454). Thousand Oaks, CA: Sage.

Stake, R. (2005). The art of case study research. Thousand Oaks, CA: Sage.

Stigler, J. W., & Perry, M. (1990). Mathematics learning in Japanese, Chinese, and American classrooms. In J. W. Stigler & R. A. Shweder (Eds.), *Cultural psychology: Essays on comparative human development* (pp. 328–353). New York: Cambridge University Press.

Stringer, E. (1999). *Action research*. Thousand Oaks, CA: Sage.

Stringer, E. (2008). *Action research in education* (3rd ed.). Upper Saddle River, NJ: Pearson.

Swanson, C. B. (2003). *Keeping count and losing count: Calculating graduation rates for all students under NCLB accountability*. Retrieved from ERIC database. (ED 480915)

Tramonte, M. (2001, April). *Risk prevention for all children/adolescents: Lessons learned from disaster intervention*. Paper presented at the Annual Convention of the National Association of School Psychologists. Washington, DC. Retrieved from ERIC Database. (ED 452462)

U.S. Department of Education. (2005a). *No Child Left Behind: A new era in education presentation*. Retrieved from http://www.ed.gov/nclb/overview/intro/presentation/index.html.

U.S. Department of Education. (2005b). *Resources + Reforms = Results*. Retrieved from http://www.ed.gov/about/offices/list/ods/resources-reform/index.html.

University of Oregon. (2004). *Dynamic indicators of basic early literacy skills. Official DIBELS Homepage*. Retrieved from http://dibels.uoregon.edu.

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

Van Manen, M (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. Albany, NY: State University of New York Press.

Wertz, F. J. (2005). Phenomenological research methods for counseling psychology. *Journal of Counseling Psychology, 52*(2), 167–177.

Whyte, W. F. (1955). *Street corner society*. Chicago: University of Chicago Press.

Wohlwend, K. E. (2009). Damsels in discourse: Girls consuming and producing identity texts through Disney princess play. *Reading Research Quarterly, 44*(1), 57–83.

Wong-Wylie, G. (2006). Narratives of developing counselor's preferred theories of counseling storied through text, metaphor and photographic images. *Qualitative Report, 11*(2), 262–301.

Xu, S., Connelly, F. M., He, M. F., & Phillon, J. (2007). Immigrant students' experiences of schooling: A narrative inquiry theoretical framework. *Journal of Curriculum Studies, 39*(4), 399–422.

Ylvisaker, M., Turkstra, L., Coehlo, C., Yorkston, K., Kennedy, M., Sohlberg, M. M., & Avery, J. (2007). Behavioral interventions for children and adults with behavior disorder after TBI: A systematic review of the evidence. *Brain Injury, 21*(8), 769–805.

APPENDIX A

Using Commercial Video Games to Improve Student Motivation

Name

Institution

Using Commercial Video Games to Improve Student Motivation

Video games have exploded in popularity in recent years. A large number of students play video games; one survey of more than 300,000 students in 50 states reported that 64% of students indicated that they played video games on a regular basis. Teachers are beginning to recognize that student fascination with video games is not going to fade away. More than 50% of teachers reported an interest in learning more about how to integrate games into their lessons and 46% said they would be willing to attend professional development sessions on the topic (Project Tomorrow, 2008). **[Paragraph provides good statistical background.]**

Video games have been studied in several different educational areas. Research has shown that commercial video games can help to improve student health; a study funded by the American Council on Exercise revealed that playing Dance Revolution (DDR) on higher difficulty settings was the exercise equivalent of riding a bicycle at 12–14 miles per hour and that it had provided similar health benefits to a high-impact aerobic workout (Anders, 2007).

When used in ways that are consistent with learning principles, video games have had positive impacts on both the achievement and motivation of students. For example, first- and second-grade students in Chile who played educational games on Gameboys in school performed better than their classmates in math, reading comprehension, and spelling. They were also more motivated to learn, less likely to act out, and more likely to interact with classmates (Rosas, Nussbaum, & Cumsille, et al. 2003). Similarly, Kepritchi, Hirumi, and Bai (2008) found that high school students who played math games in class showed higher gains in performance on district benchmark exams than their classmates. **[Brief summaries of past research studies.]**

Even games created for recreational purposes can help students learn more in school. Simpson and Clem (2008) found that the commercial game Restaurant Empire helped motivate students to learn how to use various Microsoft Office programs and that they were able to make significant connections between the game and their lives. Furthermore, Squire (2005) found that the popular Civilization III helped students who traditionally struggled in school become more interested in History, and that it encouraged them to conduct research so that they could progress in the game.

There are a few different aspects of gaming that make it an attractive addition to the classroom. Some studies suggest that games can increase students' interest in a topic to such a degree that they begin to research a topic with an interest in understanding it and not just so they can memorize facts for school (Squire & Jenkins, 2003). In addition, video games can help students to master the skills that are in demand by employers today, help them bridge the gap between theory and practice, and personalize learning (Federation of American Scientists, 2005). Adventure games in particular could be of benefit to students, as their very nature promotes problem-solving and critical thinking skills, while their use of plot hooks and emotional proximity work together to capture student interest (Dickey, 2006b on p. 459).

Despite all of the potential benefits of gaming in the classroom, there still exist some problems with using games in the classroom. For example, Sanford, Elisa, Facer, and Rudd (2006) ran into a number of problems when conducting a study on games in the classroom. Some were technical in nature, and included difficulty installing games on school computer networks, outdated and insufficient computer hardware, a lack of the necessary number of computers, and problems with saving games. Others involved the limitations on time within a school day, and teachers expecting students to be more proficient using the games than they actually were. **[Note that more than half of sources in References were cited in Introduction.]**

Statement of Purpose

While it is clear that video games have the potential to make education more effective and engaging, it is also clear that more research needs to be done on the topic, particularly in regard to the use of commercial video games. Studies have suggested that commercial video games can help to increase student motivation, but few have focused on this exclusively. **[Identifies a gap in past research.]** For this reason, this action research study will focus on the question of whether or not commercial video games can be used to increase my students' desire to study topics in which they previously showed little interest. **[Statement of purpose identifies the type of study and its purpose.]**

Review of Literature

Most of the research on the topic of video games has been conducted in the past 10 years and includes studies using survey (Sanford et al., 2006), experimental designs (Kebritchi et al., 2008; Simpson & Clem, 2008), case studies (Squire, 2005), and mixed methods research (Rosas et al., 2003). This review of literature will explore the frequency of use of video games, student health issues, the relationship among student achievement and motivation and video game use, classroom implementation, and challenges of using video games in classrooms. **[First paragraph gives an overview of types of research and identifies themes that the rest of the review will examine.]**

Video games are more popular than ever before, especially among school-aged children. As part of a yearlong study on the use of commercial video games in the classroom, Sanford et al. (2006) surveyed 2,334 students in the United Kingdom and Wales and found that 85% of children ages 11–16 played games at least once every two weeks. These students were also, in general, enthusiastic about the possibility of playing video games in school; 62% said they would like to use games in the classroom, and 55% believed that games would help to make lessons more interesting. In 2007, Project Tomorrow (2008) surveyed 319,223 K–12 students in all 50 states and reached similar conclusions. In their study, 64% of students indicated that they play video games on a regular basis. They also found that 51% of students believed that games would make it easier to understand concepts, and 50% said that games would make them more engaged in the subject matter. This popularity illustrates the need for research on the topic; if video games can, in fact, effectively help children learn, they hold the potential to revolutionize education and change the way classes are taught forever. **[Note review of literature begins with broad topic of frequency of use of video games and then moves on to discuss games in specific settings.]**

Even the Army has recognized the potential benefit of video games to influence and teach young minds. In 2002, they released a video game titled America's Army, a multiplayer online game designed to accurately represent tactical combat in a war zone and serve as a recruiting tool. Shortly after the game's release, the Army learned that officials at Fort Benning were using the game for training purposes. Recruits who failed the real-life obstacle course or rifle range were instructed to complete those levels in America's Army; after completing the levels, the recruits frequently passed the real rifle range and obstacle course tests (Zyda, 2005).

Perhaps recognizing the ability of video games to reach their students, most teachers are open to the idea of using video games in their classrooms. Sanford, Ulicsak, Facer, and Rudd (2006) surveyed 924 primary and secondary teachers in England and Wales as part of a yearlong study on the use of commercial video games in the classroom. They found that, even though only 38% of the teachers played games in their spare time, 59% of all teachers were willing to consider using them in their classrooms. When asked why they would consider using games in the classroom, 31% cited "motivating students" as the primary reason. Project Tomorrow, as part of the study mentioned earlier, surveyed 25,554 teachers, and their results were similar to the Sanford Study (Project Tomorrow 2008). For example, they discovered that 65% of teachers believed that games would help increase student engagement, that 65% believed that games would help address students' different learning styles, and that 40% believed that they would help develop students' problem-solving and critical thinking abilities. They also found that more than 50% of teachers were interested in learning more about how to integrate games into their teaching strategies, and that 46% were interested in professional development on the topic. **[Note how similar studies are grouped together above.]**

One way in which video games have been integrated into teaching is as a tool to increase student health. In recent years, the popular arcade game Dance Dance Revolution (as well as games based on it) have started to pop up in Physical Education classes. In West Virginia alone in 2007, 765 schools announced plans to integrate the game into their curriculum within two years (Anders). The game requires a lot of physical movement; the player is required to step on four directional arrows located on the ground as they light up on the screen. A study commissioned by the American Council on Exercise (Anders, 2007) showed that the game can serve as an effective form of exercise. The researchers measured the oxygen consumption, heart rate, and ratings of perceived exertion (PCE) of 24 volunteers as they ran on a treadmill. The same measures were monitored as participants completed DDR routines of various difficulties. They found that, while the lowest difficulty settings are suitable only as a warm-up, "the standard and difficult modes gave the test subjects a tremendous workout" (p. 8). More specifically, DDR was found to be comparable to riding a bicycle at 12 to 14 miles per hour and was also found to provide people with health benefits similar to those achieved during high-impact aerobics.

While increasing student health seems a natural application of video games, many studies have also explored how interactive games created for educational purposes can have a positive effect on students' academic achievement. **[Note how transition segues from previous topic, student health to the next topic, academic achievement and motivation.]** Kebritchi, Hirumi, and Bai (2008) conducted a study in which they trained teachers in the use of two single-player and three multiplayer math video games and encouraged them to use them as frequently as possible during class and computer lab time. After 18 weeks of play, the researchers concluded that students who played the games in class showed greater gains in performance on districtwide benchmark exams than their classmates. Additionally, interviews with students showed that the games had increased student motivation and that students believed the games had improved their understanding of math. Rosas et al. (2003) reached similar conclusions. In their study, they distributed Gameboys with specially designed educational games to first- and second-grade classrooms in Chile. Teachers were allowed to incorporate the games into their lessons as they saw fit, and students played for an average of 30 minutes per school day. After three months, researchers found that the students who played the games performed better in Math, Reading Comprehension, and Spelling than students who did not play at all. They additionally found, much to their surprise, that while the video games were being used, students were more motivated to learn, less likely to act out, and more likely to interact with and cooperate with their classmates. Researchers did also find, however, that control groups in the same school as the experimental groups performed similarly at the end of the experiment. They dismiss this finding as Hawthorne's effect, meaning that the control group worked harder just because they were in a study. However, this means we do not know what the true difference between the groups would have been. **[Student begins to critique methods and reasoning in studies.]**

Not all of the studies have focused on games designed strictly for educational purposes; some have examined the potential of games created primarily for entertainment. For example, Simpson and Clem (2008) integrated the commercial game Restaurant Empire into a middle school Beginning Computers class plagued by outdated materials and unmotivated students. Students were placed into teams of three and put in charge of one aspect of restaurant management. As the class progressed, students were instructed to use programs such as Microsoft Word, PowerPoint, and Excel to produce documents illustrating the progress of their businesses. They found that merely mentioning that they would be playing a video game was enough to increase the students' motivation. At the end of the class, they discovered that students made strong connections between the game and "real life." One student noted the power that the game had by saying,

> In this class my goal starting out was to get good grades like in every other class. But after a week, I started wanting to make the restaurant better, but I held that back. . . . You and your group have to manage the money, the people, food, employees, and so on. I am not sure what everybody thought about it but I thought it was all good. I learned that I do not want to be a business owner! (p. 10) **[Note how quotes of 40 words or more are indented with no quotation marks. For long quotes, the period goes before the page number at the end of the quote.]**

Squire (2005) reached a similar conclusion after integrating the commercial video game Civilization III into an urban high school and an urban middle school. In this comparative case study, he collected data for two years on how the students and teachers used the game in the classroom. He found that the game was particularly successful in reaching students who were ordinarily uninterested in Social Studies and History. Many of the students enjoyed recreating famous moments from history, while others enjoyed creating moments of revisionist history, such as playing out a scenario in which the Native Americans successfully resisted European colonization. Moreover, he found that students who failed game scenarios were forced to conduct geographic or historic research so that they could replay the scenario and progress in the game. A game scenario might, for example, ask the player to recreate the Roman Empire in the game and then prevent it from falling. A player who fails while attempting this, therefore, may be required to look up the causes of the fall of the empire in order to prevent his in-game version from collapsing. Through these failures, many of them were able to develop complex understandings of the relationship between geography and history. The British Educational Communications and Technology Agency (2006) also found that commercial games worked well to engage students, helped them improve their skills using a computer, helped improved their thinking skills, and even led them to use the library more frequently. It should be noted, however, that the researchers in the study made no attempt to control for the novelty effect; it is entirely possible that students were motivated by the idea of playing a video game in school and not by the lessons themselves. **[Good critique of limitations in study's methodologies.]** Further studies with this group of students would help to show whether or not this was the case.

The literature suggested a number of possible reasons why video games positively affect student achievement and motivation. Squire and Jenkins (2003) stated that “a good game can function as a gateway through which students can explore a much broader range of knowledge” (p. 29). That is, games can motivate students to turn to textbooks and other sources intent on understanding topics rather than simply memorizing them. In 2005, the Federation of American Scientists convened a summit featuring 100 people pulled from the Federation of American Scientists, the Entertainment Software Association, and the National Science Foundation to study video games in educational environments. They determined that video games can help students master the skills in demand by today’s employers by: bridging the gap between the learning of theories and their practical application; motivating students to set and achieve goals, even after failure; providing students with clues and hints to keep them progressing; and by personalizing learning (Federation of American Scientists, 2006). Furthermore, Dickey (2006a) believed that adventure games are particularly useful in reaching students because of their focus on problem solving and critical thinking, their narrative structure, and their use of the literary techniques of plot hooks and emotional proximity. Dickey (2006b) **[Note citation for multiple articles by same author in the same year.]** also suggested that massively multiplayer online role-playing games (MMORPGs) hold the potential to greatly benefit students, as they provide students with “intrinsic motivation” by giving them “choice, control, collaboration, challenge, and achievement” (p. 270). In addition to motivating students, MMORPGs foster learning by forcing players to think and plan critically and act strategically.

The research made a number of suggestions about how to integrate video games into the learning environment. Squire and Jenkins (2003) presented several different scenarios in which games could be used to enrich students' learning. For example, they suggested using Civilization III, a game that Squire has used in classrooms, to introduce students to historical ideas such as monarchy and monotheism, as well as to help them tie together history's different periods. They also suggested using a game known as Revolution to teach students about life during the American Revolution. The game is an online multiplayer game in which each student takes on a role, such as a loyalist blacksmith or a patriot banker. They carry out daily rituals and routines, with each student contributing something to the game environment. Such a game would be beneficial to students because:

> Role-playing games such as *Revolution* carry this process to the next level. You do not simply visit Williamsburg for an afternoon; you become part of that community. You do not simply discover what daily life was like in Colonial America; you watch the process by which the coming Revolution impinged upon and impacted those routine practices. (p. 17)

Though much of the research has been positive, not all of the news is good. **[Paragraph following provides balanced perspective by considering negative findings.]** As a part of their larger study, Sanford et al. (2006) gave 14 teachers in four different British schools access to the commercial games Roller Coaster Tycoon 3, The Sims 2, and Knights of Honor. Following a workshop designed to familiarize the teachers with the game, they each selected one game to use with their students. They were allowed to use the game when, where, and how they saw fit over the course of the 3 months. Teachers reported on their experiences and thoughts about the project through a series of interviews, e-mails, and contributions to a wiki. Additionally, teachers submitted all of their lesson plans and supporting material for analysis. Upon completion of the study, researchers, while recognizing and believing in the potential of games to help students, reported a number of challenges that must be overcome for the medium to meet its true potential. The issues they encountered included technical issues, such as installing games on district networks, running the games on computers with outdated graphics cards, and saving games for use in a future session. Other problems were institutional, such as organization of time in schools and classrooms (it often required longer to complete a task in a game than school scheduling would allow) and the availability and location of computers. Still other potential problems included the teachers' own experiences with games, teachers' expectations that children will be naturally fluent with a game from the outset, and teachers' ability to break through the game's content and use it for appropriate educational gains. Upon reviewing some of the current literature, Rice (2007) found similar barriers preventing video games from being used more regularly in the classroom. He reviewed recent research on using video games in the classroom and compiled a list of barriers that included: the perception among teachers that video games are cognitively worthless, lack of hardware sufficient to run games in schools, time divisions within the school day that do not lend themselves well to video games, and a lack of easily noticeable alignment to educational standards. He additionally found that the poor graphics (when compared to games created for fun) of educational games hindered students' desire to play and that video game environments are not yet able to provide enough interactive affordances to students.

Squire (2005) also ran into some potential problems when he attempted to use Civilization III in urban middle and high schools. Though, as mentioned earlier, he found that "underachievers" loved playing the game, he also found that more traditionally successful students hated **[Examines effects of games with diverse groups of learners.]** it and that "they were not convinced that success in a game-based unit would help them on college entrance exams or in college classrooms, both of which rely on more traditional literacies" (p. 7).

Though video games have been around for a long time now, most of the research about their use in the classroom has been done in the past 10 years. Much of this research has shown that educational video games can be used to both engage student interest and help them learn. Some research has even suggested that games produced for commercial purposes can be used to help increase student motivation and achievement. There are questions of whether the studies continued long enough to alleviate the novelty effect and researcher bias is a potential problem because most authors were advocates of using computer programs in the classroom. Many of these articles are theoretical, and although their arguments in favor of video games are based on sound learning principles, few published studies have focused on the practical use of commercial video games in educational environments. **[Note how rationale summarizes strengths and weaknesses of the literature and identifies need for proposed study.]** For this reason, this study will focus on the question: Can commercial video games be used to increase my students' motivation in subjects in which they previously displayed little interest? For the purposes of this study, a commercial video game will be defined as a game that was designed and developed primarily for entertainment and not as a classroom tool. In addition to the primary research question, several subquestions will help to guide my action research. These subquestions include:

- What are my students' attitudes toward studying the Social Studies?
- How do my students view the different learning tools used in the class?
- What changes, if any, occur in my student attitudes and motivation?
- What are my student perceptions of how video games changed what or how they learned?

Method

Participants

The participants for this study will be students from two of my sixth-grade Social Studies classes from a middle school in upstate New York. This sample of 50 students will be selected by convenience sampling; I am the teacher in both classroom settings. In addition, many of these students display a lack of interest in studying the Social Studies, which makes them ideal candidates for the study. All the students will participate in opening and closing surveys, and the results of those surveys will be used to select a smaller subsample for interviews. This allows me, the teacher-researcher, to select and speak with students who have different opinions of video games as an instructional tool. The data will be collected during class meetings, which last for approximately 40 minutes and occur every school day, from the middle of May until the end of the year. The curriculum calls for students to be studying the Eastern Hemisphere.

According to the New York State Department of Education (2006), during the 2005–06 school year, the students' school had 542 students in Grades 5–8. Approximately 35% of the students qualified for free lunch and 11% qualified for reduced-price lunch. In all, 4% of the students were limited English proficient students, 12% were African American, 24% were Hispanic or Latino, and 63% were white. In the community itself, 69.1% of the population is white, 11.7% is African American, and 28.4% is Hispanic or Latino (of any race). Approximately 68% of residents have at least a high school diploma, while 13.4% have a bachelor's degree or higher. The median household income is $27,474, and 20.8% of families are below the poverty level (U.S. Census Bureau, 2000).

As mentioned previously, many of the participating students are uninterested in Social Studies. In the past, I have tried different techniques in an attempt to spark their interest. On top of standard activities such as reading the textbook and taking exams, students have been assigned research papers, solo projects, and group projects. They have been on field trips to historic sites and to plays based on famous events and watched videos about history. They have witnessed multimedia presentations and read several different novels based on important historic events. So far, these activities have done little to generate genuine student interest in the subject.

Role of the Researcher

I have been employed as a middle school teacher in the middle school involved in the study. I am currently taking a class in educational research in which I am learning how to create and carry out the various aspects of this study. My professor will provide me with comments and feedback, helping me to refine the study as it progresses.

I have been an avid video gamer for my entire life. I have played games across nearly every genre and platform that has been released in the past 25 years. With this experience, I am highly familiar with which available games could be used to help students. I am quick to learn new games and proficient enough to solve most technical problems that could arise during play. I also have, however, some preconceived notions about the potential value of video games that could lead to bias in the study. In particular, I believe (based on my own experiences) that commercial video games, when used correctly, can be used to motivate students.

In order to help control for this, I will review all of my reflections and interview recordings with a peer debriefer, allowing me to gain an outsider's perspective on everything. Additionally, I will compare my notes with those of my teaching assistant, who will be in the room for all of the game-based lessons.

Methods of Data Collection

Several types of data collection will be used during this study to address the subquestions. Table 1 shows the subquestions and the measures used to obtain data on each one.

Students will be given a preliminary survey (see Appendix A) and a follow-up survey (see Appendix D) to assess their attitudes toward Social Studies and the learning tools used. The preliminary survey will be given in class prior to the start of the video game lessons and will be used to establish student perceptions of Social Studies at the beginning of the study. Students will be unable to anonymously

TABLE 1. Subquestions and Methods of Data Collection

Subquestion	Methods of Data Collection
What are my students' attitudes toward studying the Social Studies?	Teacher reflection journal Student surveys
How do my students view the different learning tools used in the class?	Student surveys Student interviews
What changes, if any, occur in my students' attitudes and motivation?	Student surveys Student interviews Teacher reflection journal
What are my students' perceptions of how video games changed what or how they learned?	Student surveys Student interviews

take the surveys, as the surveys will be used to select students for interviews later in the study, but their identities will be kept confidential. Both surveys will be pilot tested in another class in the school and reviewed by a second teacher to determine if items are clear and appropriate. Comments by the students and teacher in the pilot study will be used to revise the survey. In addition, the survey contains several reverse items to establish internal validity.

Following the survey, students will begin the video game lessons, using the commercial game Civilization IV (described in Appendix B). My teaching assistant and I will observe each lesson. Immediately following each lesson record, my teaching assistant and I will record notes in a reflective journal. The following prompt questions will guide each journal entry: What is the level of participation by students today? To what extent were all students engaged in the learning activities? What problems or success were evident in student interactions and comments?

My assistant and I will record their reflections separately in journals. A sample recording page for the journal is included in Appendix C. Following each lesson, my assistant and I will meet to compare and discuss our notes; this will reduce bias on my part by allowing me the opportunity to discuss what I have witnessed with someone else who was in the room. In an additional attempt to reduce bias, I will share my notes with a peer debriefer, garnering a third opinion on the information.

After the Civilization IV lessons have concluded, I will give the students a follow-up survey that is similar to the first one (see Appendix D). This survey will contain questions similar to the preliminary survey and will help me identify which students, if any, changed their opinion on Social Studies because of the game. All students who initially showed negative attitudes toward Social Studies will then be interviewed. The questions to be asked during the interviews are included in the interview protocol in Appendix E. The interviews will be tape recorded, so that I can transcribe them later. Students will then be invited to review these transcripts to ensure that their opinions are accurately reflected. After these interviews, I will once again review his materials, including the tape recorded interview sessions, with a peer debriefer to minimize bias. The use of multiple data collection tools for triangulation of results.

Procedures

The study began with my own reflections on my students' low levels of motivation while studying Social Studies, in particular the topic of the Eastern Hemisphere. After discussion of this problem with my teaching assistant, we decided to initiate the proposed action research study on the topic.

The goal of this action research study is to examine whether or not commercial video games can be used to increase my students' desire to study topics in which they previously showed little interest. The primary participants will be a sixth-grade Social Studies class made up of 25 students, myself (the teacher), and my assistant, a collaborator. The study will begin with a brief survey given to students in early May, after the bulk of the required material has been covered. This survey will allow me to gather information about my students' opinions on Social Studies.

After the survey has been completed, I will begin the Civilization IV-based lessons. The computer game requires students to take turns building an empire and trying to achieve goals such as managing a population, conquering other civilizations, or winning an election. (For more information see Appendix B.) During the game, my students will answer questions embedded in the game about how activities and decisions in the game relate to historical events in the Eastern Hemisphere.

Because the regular classroom does not have enough computers for all the students, the lessons will take place in the computer lab. These lessons will run from early May until the end of the school year. Students will have approximately 40 minutes at a time to play the game, as this is the length of each period, and will play an average of two days per week. On the other days of the week, students will read original writing on the time period, view historical videos, and engage in group discussion or the readings and videos. My assistant and I will record reflections after each class for both days on which computer games are played and days on which they are not played. Class sessions during which students are not playing the game will be used to provide information to supplement material learned or experienced in the game.

Upon completion of the game lessons, students will be given a second survey that will assess their opinions of lessons and of Social Studies. I will review the results from the surveys to select students for more in-depth interviews. Selected students will be include two types: those who initially were not interested in Social Studies but whose attitudes changed after the game, and those who showed no change in motivation. The interviews will be confidential, but they will be tape recorded in order to ensure their accuracy and so that they can be reviewed later with a peer debriefer. I will acquire parental permission to conduct the interviews and survey (see Appendix F), as some parents may not want their children to take class time to complete these measures. The interviews will be largely unstructured; the goal is to find out what students liked or did not like about playing the game and why it did or did not affect their motivation. After all data has been collected, it will be triangulated by comparing students' answers on both surveys and, if possible, the interviews.

The design of the entire study including the cycles of data collection are displayed in Figure 1, which is included in Appendix G.

Each survey will be analyzed separately. For items with a Likert scale, the mean response for each item will be calculated. A frequency polygon will be used to display the mean ratings of students' perceptions of instructional approaches. Interview data will be coded and analyzed to identify major themes.

Benefits and Limitations

In completing this study, I hope to find out whether or not students can benefit from the use of commercial video games in their classrooms. If students can, in fact, gain something from these games, it would give myself, and other teachers, another tool to increase student motivation. Even if playing a game does not increase motivation for the entire class, it is still something that could be used after school or outside of normal class time to motivate individual students. The interviews would also provide information about why some students might not like video games and might help identify the types of students who would be most or least likely to be helped by this tool. At the very least, commercial video games could provide a strategy to motivate students who dislike Social Studies; at most, they could become a tool used to generate discussion among students about different ways to learn.

One of the strengths of this study is that there will be a large amount of data collected. Comparing the different surveys, teacher reflections, and student interviews increases the validity of the study. This allows me to speak to a specific student if I notice a difference in his answers on the surveys and his behavior in class. Additionally, strong attempts are being made to limit the amount of bias; everything collected and recorded will be reviewed with a peer debriefer, and classroom observations and notes will be compared and discussed with my assistant in the room.

There are several possible limitations to this study. Despite all the attempts made to control it, it is still possible for some of my bias to shine through. Another potential problem could be the novelty effect; some students may experience benefits from the game simply because it is something new and different. This allows for a possible future study in which students who reported benefits from the game are studied down the line to see if repeated game play maintained those benefits. Finally, the study could be skewed somewhat by the Hawthorne effect. Some students, aware that they are being studied, may alter their answers on surveys and in interviews in order to please me. This is one reason that my assistant and I will take detailed notes as students are playing the game; it would be difficult for them to feign interest throughout the entire course of the study.

VIDEO GAMES 24

References

Anders, M. (2007). Human joysticks. *Fitness Matters, 13*(5), 7–9.

British Educational Communications and Technology Agency. (2006). *Computer games in education project: Report.* Retrieved from http://partners.becta.org.uk/index.php?section=rh&catcode=&rid=13595&pagenum=1&nextstart=1

Civilization IV. (n.d.). In *Wikipedia*. Retrieved from: http://en.wikipedia.org/wiki/Civilization_iv

Dickey, M. (2006a). Game design narrative for learning: Appropriating adventure game design, narrative devices and techniques for the design of interactive learning environments. *Educational Technology Research and Development, 54*(3), 245–263.

Dickey, M. (2006b). Game design and learning: A conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation. *Educational Technology Research and Development, 55,* 253–273. doi:10.1007/s11423-006-9004-7

Federation of American Scientists. (2006). *Harnessing the power of video games for learning.* Retrieved from http://fas.org/gamesummit/resources/summit%20on%20educational%20games.pdf

Kebritchi, M., Hirumi, A., & Bai, H. (2008). *The effects of modern math computer games on learners' math achievement and math course motivation in a public high school setting.* Retrieved from http://www.dimensionm.com/docs/ucfresearch_brief_june_202008.pdf

New York State Department of Education. (2006). *The New York State School Report Card:*

Ellenville, NY. Retrieved from http://www.nystart.gov/publicweb-rc/2006/AOR-2006-622002060005.pdf

Project Tomorrow. (2008). *21st century students deserve a 21st century education.* Retrieved from http://www.tomorrow.org/docs/National%20Findings%20Speak%20Up%202007.pdf.

Rice, J. (2007). New media resistance: Barriers to implementation of computer video games in the classroom. *Journal of Educational Media and Hypermedia, 16*(3), 249–261.

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., et al. (2003). Beyond Nintendo: Design and assessment of educational video games for first and second grade students. *Computers and Education, 40*, 71–94.

Sanford, R., Ulicsak, M., Facer, K., & Rudd, T. (2006). *Teaching with games: Using commercial off-the-shelf computer games in formal education.* Retrieved from http://www.futurelab.org.uk/resources/documents/project_reports/teaching_with_games/TWG_report.pdf

Simpson, E., & Clem, F. (2008). Video games in the middle school classroom. *Middle School Journal, 39*(4), 4–11.

Squire, K. (2005). *Changing the game: What happens when video games enter the classroom?* Retrieved from http://website.education.wisc.edu/kdsquire/manuscripts/innovate.df.pdf

Squire, K., & Jenkins, H. (2003). Harnessing the power of games in education. *Insight, 3*(1), 5–33.

U.S. Census Bureau. (2000). *Fact sheet: Ellenville, NY.* Retrieved from http://factfinder.census.gov

Zyda, M. (2005). *From visual simulation to virtual reality to games.* Retrieved from USC Information Sciences Institute Web Site at: http://gamepipe.usc.edu/~zyda/pubs/Zyda-IEEE-Computer-Sept2005.pdf

Appendix A

Name ______________________________

Preliminary Survey

You are taking part in a study on the use of different approaches to learning about Social Studies. Please answer all the following questions honestly. Your participation in this study is voluntary and not participating will have no effect on your grade. If you do participate, nothing that you say on this survey will have any effect on your grade. Your name and your responses will also not be shared with anyone else. Please return the survey to [teacher researcher name] upon completion. Results from the study will be shared with all students in class at the end of the semester.

Directions: Read the questions below and respond by circling the number that best corresponds with how you feel using the scale below.

1 = strongly disagree

2 = disagree

3 = slightly disagree

4 = slightly agree

5 = agree

6 = strongly agree

VIDEO GAMES 27

Attitudes Toward Social Studies

I enjoy studying history	1 2 3 4 5 6
I enjoy studying geography	1 2 3 4 5 6
I find Social Studies class to be boring	1 2 3 4 5 6
I hate studying geography	1 2 3 4 5 6

Student Perceptions of Instructional Approaches

Directions: Indicate how you feel about following instructional approaches by writing a number from 1 to 5 to indicate which methods you like the most and in which approach you feel you learn the most. The number 1 would indicate that you like or learn the most from a method and 5 would indicate that you like or learn the least from a method.

Method	**Like (1= most; 5 = least)**	**Learn (1 = most; 5 = least)**
Reading books or articles		
Class discussion		
Videos		
Computer games		
Lectures		

Appendix B

Information on Civilization IV

Civilization IV is a turn-based game in which the player builds an empire from scratch. All standard full-length games begin in 4000 BC with a settler who builds a single city. From there, the player expands an empire while contending with rival nations, using the geography, developing infrastructure, and encouraging scientific and cultural progress. By default, players can win the game by accomplishing one of five goals: conquering all other civilizations, controlling the majority of the world's land and population, being the first to construct and launch a space ship capable of colonizing Alpha Centauri and have it land there first, increasing the Culture ratings of three different cities to "legendary" levels, or by being declared "World Leader" by winning a popularity election through the United Nations. If the game's clock runs out (by default in the year 2050 AD) with none of these goals fulfilled by any nation, the nation with the highest score is declared the winner ("Civilization IV").

Appendix C

Reflective Journal Entry

What is the level of participation by students today?

No or few students actively involved	Most students actively involved	All students actively involved
1	2	3

To what extent did students appear interested in the Social Studies topics?

Little or no interest	Moderate amount of interest	Deep level of interest
1	2	3

What problems or success were evident in student interactions and comments?

Appendix D

Follow-Up Survey

Please answer all of the following questions honestly. Nothing that you say on this survey will have any effect on your grade, and your name will not be shared with anyone else.

Directions: Read the questions below and respond by circling the number that best corresponds with how you feel.

1 = strongly disagree; 2 = disagree; 3 = slightly disagree;
4 = slightly agree; 5 = agree; 6 = strongly agree

I enjoyed playing the game in class	1 2 3 4 5 6
I learned something about Social studies that I did not previously know by playing the game	1 2 3 4 5 6
I enjoyed Social Studies more during the time that we played the game	1 2 3 4 5 6
I found the game to be boring	1 2 3 4 5 6
I preferred noncomputer-based Social Studies classes to playing the game	1 2 3 4 5 6

Please write a short answer to the following questions

After playing the game, did you use the Internet in order to do research? If so, please describe what you looked up and why.

VIDEO GAMES 31

What if anything, did you learn anything by playing the game? Please briefly describe anything that you learned.

Appendix E

The Interview

This is a general outline of what questions could be asked during the final interview. The actual interview may contain different questions, depending on the student and information gathered from the reflective journals. Prior to the interview, I will explain that these interviews are a part of a study on using different learning approaches in the classroom, and that they were selected based on their answers on the survey and based on teacher observations. They will be informed that they can stop their participation at any time and may decline to answer any question. Additionally, they will be asked to answer honestly and reassured that their answers will not have any effect on their grades.

- Tell me about your experience in playing the computer game. What about it did you like and dislike about playing the game in class?
- Did the game ever get you to think about something you had not previously considered? If yes, explain.
- How would you compare the game-based lessons to a nongame based classroom lesson? Which do you prefer? Why?
- Did your attitudes toward Social Studies change or not change while playing the game? If it changed, in what way did it change?
- Would you play the game in your spare time? Why or why not?

Appendix F

Parental Permission for Interviews

As your child's teacher I plan to conduct a simple action research study designed to improve my students' attitudes toward Social Studies. Specifically, I will be examining several learning strategies for teaching Social Studies. These will include use of videos and video games, historical books, and group discussions. As part of the study students will complete two brief surveys on their attitudes toward social studies. Several students will also be interviewed about their feelings about the learning strategies used. All information will be kept confidential and participation is voluntary. No child is required to participate and whether they participate or not will in no way affect their grade. Please indicate if your child can take part in the study by signing the statement below.

I give permission for my child, ______________________________________, to take the surveys and to be interviewed as a part of a study on different learning strategies for Social Studies in the classroom. I understand that the interview sessions will be tape recorded, but that my child's name will not appear anywhere in the final report, and that his or her name will not be available to anybody but the teacher-researcher. I also understand that all information from the surveys will be kept confidential.

______________________________ ____________________

Guardian Signature Date

APPENDIX B

How Do Literacy Teachers Describe Their Strategies for Teaching English Language Arts

Name

Institution

How Do Literacy Teachers Describe Their Strategies for Teaching English Language Arts

For years professionals and practitioners in the education field have dealt with the debate of whether to teach English Language Arts (ELA) using phonics-based instruction or whole language-based instruction. Throughout this discussion some teachers have been mandated by their schools to use a specific method, while others have tried to blend the methods. While there is research support for both methods, most current studies suggest a blending of both approaches.

O'Donnell and Wood (2004) define phonics instruction as teaching a specific set of subskills preceding learning how to read. To learn these subskills, O'Donnell and Wood suggest that children engage in extensive practice with phonemic segmentation activities, synthetic phonic exercises, and work with decodable text before they try to read real world texts. One example of phonics instruction can be seen in the Basal Reading Series. These series are a collection of stories accompanied by workbooks and skill sheets; these practice materials focus on a skills hierarchy explicitly teaching word identification skills and sight word identification, phonics, and rules for reading.

Whole language instruction emphasizes meaning. Whole language believers work with the premise that reading and writing is most naturally learned from whole to part, in a meaningful context. O'Donnell and Wood (2004) suggest that a teacher working with the whole language model would provide students with a print-rich environment, in which they are encouraged to read, write, and explore books, and are given the opportunity to observe reading and writing in real life. Realistic but predictable texts are used in a whole language classroom allowing teachers to include direct teaching of skills or concepts in a natural, functional way rather than being isolated and in a prescribed sequence as phonics instruction suggests.

The debate on whole language vs. phonics has produced a huge amount of literature including numerous empirical studies. There is evidence, however, that both methods work when applied appropriately as separate instructional approaches. Other research suggests that these methods should be used together, in a blended method, in order to produce the best results for students.

Phonics-based instruction is supported by research completed by Kamps et al. (2007), who examined how a phonics-based intervention affected English Language Learners' progress in reading. This study found that the participants improved their early literacy skills using the phonics-based program. Buck, Treiman, Caravolas, Genesee, and Cassar (1998) provided even stronger support for phonics-based approaches in a study with two groups, one group receiving phonics instruction and one group receiving whole language instruction. The researchers concluded that while both groups were equally knowledgeable about the basics of spelling, the students receiving phonics instruction spelled more age and grade appropriately than the students receiving whole language instruction.

While the previously stated research supports phonics instruction there is also research that favors using a whole language method in the classroom. Manning and Kamii (2000) compared two groups receiving either whole language or phonics instruction. They found that even though the whole language group began the year at lower levels, they made more progress in reading and writing, developed coherence, and had a greater absence of regression and confusion when compared to the phonics group.

While the research produces conflicting results regarding which approach is best, a close analysis of the studies shows that different measures (dependent variables) were often used. For example, phonics is superior when spelling is the measure and whole language students perform better in reading and writing. More recent research has supported a blended method, using both phonics and whole language instruction since all of these components of reading success are important. LeDoux (2007) found that teachers believed that whole language method seemed to create a love of literature, foster problem solving and critical thinking, encourage collaboration, and provide authenticity. However, they argued that it could not be used exclusively. Whole language instruction should include an element of phonics and decoding in order to be the most effective. Quantitative research also supports that an approach blending whole language and phonics produces the greatest gains in reading comprehension (Reutzel, Fawson, & Smith, 2008).

While there is extensive research supporting the use of whole language, phonics, or a blended method, there seem to be few studies that examine what real practitioners in schools that are meeting state standards are actually using. Studying schools that are meeting state standards may shed light on what teachers believe are effective ELA practices.

Statement of Purpose

Therefore, this study proposes to describe what practices and approaches primary school literacy teachers, whose students are meeting state standards, use in order to teach ELA. Areas that will be investigated will include curriculum and assessment ideas, instructional strategies, and theoretical orientations.

Review of Literature

The discussion of how to best use whole language and phonics instruction continues among practitioners and researchers in the field of literacy. While the goal of all practitioners is to provide their students with the best possible education, beliefs in one method over another will influence teaching and learning. A review of current research on the best way to teach literacy to students has yielded few findings that consistently support one method over another while the majority of the research supports blended whole language and phonics instruction. Most research in this area has been quantitative, using standardized test scores to represent improvement in achievement. Some research has used a mixed method approach; however this research is limited. Literature on this topic is organized into categories that focus on one of three areas: phonics instruction, whole language instruction, or a blended instruction.

Before discussing studies that favor phonics instruction, whole language instruction or a blended approach, it should be noted that some studies find no differences in reading ability when comparing these two instructional approaches. For example, a very recent study explored different methods of teaching literacy that were neither defined as whole language or as phonics instruction, but under the surface used elements of both instructional methods. In this study a group of 72 third-grade students were given a series of pre-tests and post-tests to assess fluency and comprehension. These assessments included the DIBELS and an Oral Reading Fluency test. The students were randomly assigned into two groups receiving instruction in scaffolded silent reading (ScSR) or guided repeated oral reading (GROR). The instructional methods of ScSR had a base in whole language instruction. Specifically the use of teacher models, conferencing, and goal setting with students, in conjunction with the use of authentic texts and immersing students in literature throughout the day are all aspects that define whole language instruction. There were many aspects of phonics instruction guiding the GROR instructional method. For example, the use of choral reading, the use of direct instruction from a text for an assessment, and the reading of one common book in the classroom are all aspects that help define phonics instruction. The findings suggested that there was no significant difference between the groups in relationship to literacy growth. Both treatments improved reading rates (Reutzel, Fawson, & Smith, 2008). While the participants were randomly assigned to the treatment groups, the differences in instruction between the groups were not distinct enough to reach conclusions about phonics versus whole language.

Similar findings were reported by Reutzel et al. (2008) and Kamps et al. (2007), whose studies showed that growth could be made using both instructional methods. The Kamps et al., study was unique in that it described secondary-tier interventions in schools serving English Language Learners. Participants in this study were selected randomly and studied over a five-year span with the goal of describing evidence-based interventions and outcomes in these schools. The participants were split into two groups, one being at risk for reading failure and one in which the students were not at risk and therefore did not receive the intervention discussed. A direct instruction approach was used in the experimental group focusing on specific skills with a focus primarily on phonemic awareness. As the students gained these skills they were moved into a balanced literacy group, including phonics-based components of guided reading. The DIBELS test and two subtests to rate oral reading fluency and nonsense word fluency were used as benchmarks and post assessments. Assessment scores showed that a significant difference was found in nonsense word fluency between the experimental and comparison groups, the experimental groups making larger gains. However, there was no significant difference between the groups in oral reading fluency suggesting no differences in this important reading subskill.

While the above studies suggest no real outcome differences in the use of phonics instruction vs. whole language instruction, for some researchers and practitioners phonics instruction is considered to be a key component to produce successful readers. For example in a study conducted by Brackemyer, Fuca, and Suarez (2001) students in kindergarten and second-grade classes received a phonics intervention in effort to increase failing reading scores. First, however, the researchers administered a survey to all grade level teams to uncover the main instructional strategies being used to teach reading. These surveys revealed that the majority of the teachers were not including daily direct phonics instruction into the literacy instruction, and felt that there was not sufficient time allotted to accommodate all necessary instructional components. Baseline data showed a very high number of students with low reading scores. In addition to lack of time and neglect of phonics instruction, the teachers were also faced with a multitude of reasons their students were failing and felt they had little help in identifying students with reading problems. Because of the two different schools of thought, whole language instruction vs. phonics instruction, they had difficulty deciphering how to help these students. The researchers proposed a phonics instruction intervention that would increase student ability to recognize beginning consonant sounds. This intervention included materials that emphasized beginning sounds, daily instruction, and extension activities. The intervention itself took place twice a week and used instruction methods such as Basal readers. As a measure of assessment pretests and posttests were given and covered the skills necessary for reading at each grade level. According to the assessments given significant gains in all reading areas were made, bringing the majority of students up to or above grade level in reading scores. The researchers believed that repeated exposure to direct phonics instruction may have helped to develop a better understanding and transfer of skills into reading. In addition to their findings the researchers also suggested that repeated exposure to phonics was absolutely necessary in any reading program in order to produce successful readers. While the study certainly produced positive findings for phonics instruction, the survey prior to the intervention revealed the lack of a cohesive literacy instruction approach among the teachers. It might be that the intervention simply provided a cohesive and coherent program which resulted in improved test scores.

Bruck, Treiman, Caravolas, Genesee, and Cassar (1998) also suggest that phonics instruction is necessary to produce successful readers. In their study the researchers explored whether a phonics method or a whole language method of instruction would produce better spelling scores. The researchers selected 54 third-grade students who had been instructed using a strictly whole language approach with absolutely no direct phonics or spelling instruction and a second group of 22 students who were instructed using a strictly phonics approach to literacy. As an assessment to measure the students' spelling ability they were asked to spell 50 words and non-words with varying phonemic and orthographic structures taken from the Woodcock Reading Mastery Test-Revised. Within a span of one week the students were asked to spell the words on the list two different times. The results of their research indicated that the whole language group was scoring significantly lower on word recognition, than the phonics group. On average the phonics group was scoring in the 60th percentile, while the whole language group scored in the 47th percentile. The whole language group was spelling inappropriately for their age and grade level. The phonics group spelled more words correctly than the whole language group in terms of using more conventional phoneme-grapheme correspondences, with the one exception that the types of spelling errors made did not differ. It has been stated that both groups were equally knowledgeable about the basics of spelling; however, the phonics group knew how to apply and use the basics with their phonemic knowledge in order to produce correct spellings. Bruck et al. concluded that as compared to children in phonics-based programs, children in whole language programs did not make the same progress on basics skills and spelling ability, suggesting that phonics was the more important instructional method to use when teaching students the basics of spelling. It is important to note that the dependent variable in this study was spelling. It would have been interesting to see if the group differences existed when comprehension or reading ability or fluency was measured.

While there is certainly research that supports and focuses on phonics instruction in the classroom to produce proficient readers, some research also suggests that using a whole language approach to literacy instruction yields better and more significant results than phonics instruction. For example, Manning and Kamii (2000) conducted research with 38 kindergarten students to determine whether whole language or phonics instruction produced better scores and more growth among students' reading and writing. The students were grouped within their classrooms, with one teacher identifying herself as a whole language teacher and the other identifying herself as a phonics teacher. The whole language class spread instruction out over the course of the day including read aloud for a total period of one hour a day. The phonics class was instructed during teacher judged appropriate moments within context; some activities included journal writing, shared reading, and modeled writing. The students in both groups were interviewed five times throughout the year using a two part process: one part focusing on reading and one part on writing, except for the first interview during which the students were only given the writing portion of the interview. The writing sections aimed to evaluate whether students wrote longer words with more unconventional letters and if the students wrote the same letters unconventionally for parts of words that sounded the same. The reading section of the assessment focused on whether the students recognized the words that were being read as actually being printed on the paper and where they perceived that word to be within the sentence. Results of the assessment and instructional procedures indicated that the whole language group started the year at lower levels however, they made more progress in both reading and writing in terms of development, coherence, and demonstrated absence of regression and confusion. In the whole language group there were only two instances of regression while there were 15 instances of regression in the phonics group throughout the year. The researchers believed that further investigation was warranted in tracking the students from the phonics group in future reading success (Manning & Kamii, 2000).

While many researchers argue that one method is better than the other, the majority of research on this topic suggests that one method cannot be used without the other. In order to truly produce successful readers and writers a literacy program must incorporate components of each method. For example, research conducted by Dahl and Scharer (2000) explored the teaching of phonics within first-grade whole language classrooms. There were five dimensions that these classrooms all had in common. They were child centered, where reading and writing were taught as meaning centered. They used a wide variety of reading materials that included the students' own print materials. The classroom created a literate environment where reading and writing were used as tools for learning and problem solving. Finally, children were encouraged to collaborate as they worked on reading and writing projects. Even though these classrooms were identified as whole language, about one third of instructional actions represented phonics instruction, including phonological awareness, phonemic awareness, and phonemic segmentation. The teachers also identified four particular strategies frequently used: making use of what the learners already knew in order to decode or encode new words, stretching out the sounds of words, rereading and rethinking miscues, and rechecking a written word to be sure it represents every sound heard. Phonics instruction was woven into daily whole language activities. Summarized findings indicated that phonics instruction was spread across a variety of standard whole language activities; writing experiences served as an essential context for developing phonics knowledge; and phonics was taught at the point of use as teachers and students engaged in reading and writing. Using pretest and posttest data in conjunction with the findings previously stated, the researchers found that these students made significant gains in literacy achievement in terms of decoding unknown words and dictation. Due to the nature of this study there was no control group used to compare data or methods. The researchers focused on individual gains of the students within these classrooms and whether they were reading at grade level in order to examine if the methods the teachers were using were effective. The researchers argued their research and similar research was working toward ending the debate that places whole language instruction and phonics instruction on separate ends of the continuum. This research demonstrated that phonics could effectively be taught in whole language classrooms and produce results that any teacher would want to see (Dahl & Scharer).

In addition to research conducted by Dahl and Scharer (2000), Joslin (1994) conducted a study that compared the effects of a pure and modified whole language instruction on decoding skills. Twenty kindergarteners were assessed using the Darrell Morris Early Reading Screening Instrument, which measures alphabet knowledge, phonological awareness, and word recognition in context and in isolation. The students were then divided into two groups of 10, each group with a range of ability levels, and both groups were placed in classrooms that were similar in physical design. Joslin hypothesized that the modified whole language group would show higher decoding automaticity than the pure whole language group. In the pure whole language group, the students received no whole class phonics instruction. A teacher statement from this group indicated that sounding out unknown words was the least effective way to figure the word out. The phonetic rules were left to be discovered. The modified whole language group received 10 to 20 minutes of focused, intensive, and systematic phonics instruction daily.

This group was not only encouraged to use phonics-based decoding skills but to use context-based skills as well. Joslin reported that his research yielded similar scores between the groups. Both groups demonstrated a firm grasp of alphabet knowledge, as would be expected by the end of kindergarten, with a slightly higher percentage in the modified whole language group. The modified whole language group yielded slightly higher scores than the pure whole language group in phonological awareness, possibly because of the familiarity of the activity in the modified group. Joslin's research also found that the modified whole language group made significantly greater gains than the pure whole language group in terms of word recognition in context and in word recognition in isolation. These findings supported Joslin's hypothesis that the modified whole language approach would have a greater effect on student decoding automaticity than the pure whole language approach.

LeDoux (2007) also conducted an interview with primary teachers that supported the use of whole language in conjunction with phonics instruction. In her study LeDoux examined the whole language approach to teaching literacy and explored its strengths and weaknesses. LeDoux hypothesized that whole language was an effective method for teaching students to read. Five experts in the field of reading were interviewed about their beliefs of best practice approaches. Three of the reading experts were practicing elementary school teachers who have since become college professors of early literacy; one was a reading specialist and one was a special education teacher. A major strength that the experts identified was that the whole language approach created motivation and interest in reading. The experts also identified that whole language was connected to the students and therefore helped them to feel successful. The experts also reported high test scores for their students which they attributed to the high-quality literature and authentic experiences used in whole language. However, while all these strengths were listed, the experts in this study agreed that whole language could not be effectively used without a component of phonics instruction. One expert stated at even in its purest form, it was always intended that if a child could not read a word using a whole language approach, phonics/decoding skills would be used as a last resort. All of the experts interviewed felt that direct and explicit phonics and decoding instruction should be included as part of the whole language instructional approach. They stated that when whole language classrooms implemented a balanced literacy instruction including high quality literature, motivation, and a phonics/decoding component, the programs tended to create successful readers who embodied a love of reading. A pure whole language program was not found strong enough to stand alone.

Further support for a blended approach to literacy was found in a comprehensive review of literature. Smith (2003) reviewed many pieces of literature that supported a balanced approach to literacy instruction. For example, Baumann and Ivey (as cited in Smith, 2003) and Pressley (as cited in Smith, 2003) found that immersion in literature when combined with embedded strategy instruction created an interdependent relationship in which each program contributed to and fed off of the other. It was found that the use of high-quality literature enhanced students' reading and fluency rates and helped to develop literacy abilities.

In the same review, Denton (as cited in Smith, 2003) found that neither "drill-and-kill" phonics instruction nor a pure whole language approach can meet the needs of all students. Research cited in Denton's study was supported by the State Board of Education and Department of Public Instruction in North Carolina. This study and its supporters indicated that phonics cannot stand alone and a balanced approach to literacy instruction is the most effective and efficient method to improve word recognition, writing and spelling activities, and reading meaningful text. Similar research was conducted by Lavadenz (as cited in Smith, 2003) and Mucelli (as cited in Smith, 2003) that also supported a balanced approach to literacy as most effective and efficient. Lavadenz found that fundamental literacy instruction used authentic materials and writing activities that included any type of literature that was meaningful to a student. Lavadenz also found that direct phonics and decoding skills lessons that were taught would be useful in not only reading success but in the real world as well. Mucelli supported teaching practices from the Reading Recovery program. Components of this program include a balanced approach to phonics and phonemic awareness instruction as well as encouraging improving comprehension and the independent reading of quality books. As clearly demonstrated through Smith's (2003) review, there are many researchers and professionals who believe a balanced approach to literacy instruction is the best most efficient approach.

The research clearly considers a blended approach to literacy as the most effective approach; however the question remains what do teachers do in real classrooms? Some limited research has explored teacher practice in a direct fashion by asking practitioners to comment about and identify their literacy instruction practices. For example, Zalud, Richardson, and Richardson (1992) conducted research that surveyed reading materials, methods, and teacher theoretical orientation in elementary and middle schools in South Dakota. The purpose of this study was to establish the extent to which certain programs, materials, and practices were being utilized in these schools. Of 248 districts that were sent these surveys, 117 (48%) were returned and 77 schools (31%) completed both parts of the survey. Part one of the survey included information on choices the teachers made among reading programs, teaching methods, standardized tests, and informal tests; part two provided the teachers with questions that would determine his or her theoretical orientation. These surveys indicated that most teachers were using structured phonics programs; only two programs were not phonics/linguistically based. Two conclusions were drawn from the first part of this survey: first, that teachers were using a systematic approach to teaching literacy, and, second, that programs that were being utilized were usually used in conjunction with other curriculum. This section of the survey also explored which Basal series were being used in the schools. The researchers found that the majority of schools surveyed were using Houghton Mifflin and Company, Incorporated, or Scott Foresman and Company. The second part of the survey indicated that most teachers, about 96%, had a skills-based or phonics theoretical belief, while only 4% believed in a whole language approach. The researchers concluded that this was in line with previous research. The overall conclusion drawn from this study was that while most teachers adhered to phonics or skills-based theory there was a holistic instruction component incorporated, knowingly or unknowingly, by the materials they chose to use in their classrooms.

While there has been a vast amount of research presented comparing the effectiveness of literacy programs, there appears to be a gap in investigating what methods teachers are actually using in schools. It would be particularly enlightening to examine the literacy methods used by teachers whose students are meeting state standards. This proposed study offers to begin closing this gap in the research by describing what practices and approaches primary school literacy teachers, whose students are meeting state standards, use in order to teach ELA. The following subquestions will guide study:

1. What theoretical orientations to teaching literacy are held by teachers whose students are meeting state standards in ELA?
2. What materials are used by these teachers during their instruction?
3. What assessments do these teachers use to measure growth?

Method

Participants

This study will involve primary grade teachers of ELA/Literacy within the Capital Region of upstate New York whose students are meeting state standards in Language Arts. The school districts will be chosen by the following criterion: 80% of the students are performing at or above a Level Three according to New York State Report Cards in ELA/Literacy; and the districts will have produced students meeting these standards for five or more years. The socioeconomic status of the districts, ethnicity of teachers, and gender of teachers will not influence the choice of participants in this study.

A cluster random sample will be used to select districts for the study. School reports cards will be used to identify school districts that meet the criteria outlined above. Four districts will be randomly selected from this population. The superintendents of the districts will be sent information regarding the purpose of the study and letters of support will be requested. Once letters of support are obtained, the researcher will meet with the principals of all of the elementary schools within the districts to solicit their support and participation. The schools will be asked to identify all of the teachers in their buildings involved in teaching ELA. These teachers will be sent a consent form via their school mailboxes. This consent form will indicate that participation is voluntary and all information will be confidential. The consent form can be found in Appendix A. When the consent form is properly filled out and returned to the researcher, the teachers will be given the survey, found in Appendix B, and asked to complete it indicating their literacy teaching strategies.

Instruments

The methods used by teachers instructing students in ELA/Literacy meeting state standards will be measured by a descriptive survey developed by the researcher. The survey will be administered electronically which will enable the participants to receive a coupon to a local bookstore in thanks for completing the survey. This survey is designed to describe best practices and approaches used by teachers teaching ELA/Literacy. The survey will include items addressing the following three areas: theoretical orientation of teachers, description of materials used in literacy instruction, and assessments used in order to measure growth in literacy. The teachers will be asked to use a checklist and the Likert scale to respond to items in order to complete the survey. A copy can be found in Appendix B.

This instrument is created to pinpoint specific techniques used by teachers whose students are meeting state standards. Using currently employed teachers will provide information about instructional methods that will be seen as attainable and feasible to other teachers in similar classrooms.

In order to achieve reliability and validity, this descriptive study will undergo a pilot test. A group of teachers as similar as possible to the group that will be surveyed will take the survey and judge whether items are clear and appropriate to assess reliability and content validity. In addition, the pilot group will be asked to comment on and assess the survey's internal consistency by identifying any items out of place. The survey will also contain reverse items to be sure the teachers are responding to items consistently.

Procedure

The researcher began this study by investigating already completed research on best practices for teaching ELA and Literacy. This literature review produced findings that were ambiguous in terms of actually being related to procedures that are currently being used in classrooms meeting state standards. This research was lacking in describing what current teachers are doing in their successful classrooms. This study aims to begin closing the gap in this research by beginning to examine what practices successful teachers are using in their classrooms; to do so a descriptive survey, found in Appendix A, was developed.

The school districts will be chosen using data obtained by the New York State Report Card system and a cluster random selection process. The districts that are randomly selected will have shown 80% of their students obtaining a Level Three or higher in ELA/Literacy in the past five consecutive years. The researcher will meet with the superintendents to discuss the study and gain permission to contact the school principals, citing success in ELA/Literacy instruction. When the districts have been identified the researcher will obtain information about the number of elementary schools found within the districts and how many elementary teachers those districts encompass.

The researchers will then contact the principals and obtain permission to send the consent form to the teachers via school mail. Once the consent forms are returned, the prospective teachers to be included in this study are identified the researcher will send the survey electronically to the teachers, with a link for a coupon to a local bookstore upon completion of the survey. The participants will be asked to return the survey within one month and the researcher will send an e-mail after two weeks to remind teachers to complete the survey.

The researcher does not anticipate any ethical issues presented through the administration of this survey. There are no experimental groups, control groups, interventions, or any other practices that may hinder a participant's growth or well-being. The participants are only being asked to report on their best practices in order to help other teachers improve their own practices.

The researcher will report findings from the Likert scale using percentages of reported responses and identifying the frequency of methods, materials, and assessments used in these classrooms. Themes from the comment section that is available at the end of the descriptive survey will be reported in a narrative form. The results will be presented in such a manner to ensure anonymity and confidentiality.

Benefits and Limitations

While there is vast research addressing phonics and whole language approaches to teaching ELA and Literacy there is a gap in this research in investigating what methods teachers are actually using in schools. This descriptive survey will enable other literacy and ELA teachers to learn about methods that teachers in the classroom are using with their students to yield gains in these areas and meet state standards.

The research methods used in this study are strong. Extensive literature reviews were completed and literature was used to inform the design of this study. A cluster random sample is used to ensure sample is representative of schools in the population. Reliability and validity will be determined by using a pilot test group and reverse questions. It is expected that results yielded from this study will increase knowledge about the practices of current teachers and whether the blend between phonics and whole language instruction is actually occurring.

The knowledge revealed from this survey will expand the literature on the actual practices used to teach reading to primary grade students. It will add information about the theoretical orientations held by practicing teachers and what materials and assessments they use.

One problem seen with this study is the sample used will not allow generalizing findings to a larger population. This survey will only speak to the methods used in four school districts in upstate New York and therefore the findings cannot be considered true for all schools receiving scores in Level Three or above in ELA and Literacy. As with all surveys return rate is also a possible weakness. It is anticipated that 70% of surveys will be returned so that the results will be considered strong. One last possible problem with this survey is that the results may not lead to a conclusive best practice. The debate could likely continue with more research showing that both practices are useful. It is hoped, however, that this survey will provide greater insight into methods that are working for students. For future research in this area it is suggested to researchers that they use a larger sample population to be able to further generalize findings.

References

Brackemyer, J., Fuca, D., & Suarez, K. (2001, December). *Improving reading skills through phonics instruction in the primary grades* . Chicago, IL. Retrieved from ERIC database. (ED459448)

Bruck, M., Treimean, R., Caravolas, M., Genesee, F., & Cassar, M. (1998). Spelling skills of children in whole language and phonics classrooms. *Applied Psycholinguistics, 19*, 669–684.

Dahl, K., & Scharer, P. (2000). Phonics teaching and learning in whole language classrooms: New evidence from research. *The Reading Teacher, 53*(7), 584–594.

Joslin, H. (1994). *A comparison of the effects of pure and modified whole language instruction the decoding skills of kindergarteners.* Albarmarle, VA: Phonological Awareness. Retrieved from ERIC database. (ED372367)

Kamps, D., Abbott, M., Greenwood, C., Mayer, C., Wills, H., Longstaff, J., et al. (2007). Use of evidence-based, small group reading instruction for English language learners in elementary grades: Secondary-tier intervention. *Learning Disability Quarterly, 30*, 153–168.

LeDoux, A. (2007, May). *Investigating the implementation of whole language: Strengths and weaknesses.* San Rafael: Dominican University of California, School of Education. Retrieved from the ERIC database. (ED496339)

Manning, M. & Kamii, C. (2000). Whole language vs. isolated phonics instruction: A longitudinal study in kindergarten with reading and writing tasks. *Journal of Research in Childhood Education, 15*(1), 53–65.

O'Donnell, M., & Wood, M. (2004). *Becoming a reader: A developmental approach to reading instruction.* New York: Pearson.

Reutzel, D., Fawson, P., Smith, J. (2008). Reconsidering silent sustained reading: An exploratory study of scaffolded silent reading. *The Journal of Educational Research, 102*(1), 37–58.

Smith, C. (2003, September). *The use of balanced instruction in language arts curriculum.* Bloomington, Indiana: Eric Clearinghouse on Reading, English, and Communication. Retrieved from ERIC database. (ED480636)

Zalud, G., Richardson, M., & Richardson, J. (1992, 12). *A survey of reading materials, methods, and teacher theoretical orientation in South Dakota public, non-public, and Bureau of Indian Affairs schools.* Paper presented at the University of South Dakota, Vermillion, SD. Retrieved from ERIC database. (ED351663)

Appendix A

Consent to Participate in Descriptive Survey Research

You are being asked to take part in a research study conducted in your school district. The purpose of this research is to identify and describe effective ELA/Literacy instructional practices of teachers whose students are meeting state standards.

Ms. [researcher's name] a graduate student from the College of Saint Rose has created the descriptive survey to be completed in effort to identify and describe these literacy instruction teaching practices. This survey is being sent to primary teachers within your school district. Your district was selected because over 80% of students in the district received scores of three or better on the state assessment. The information derived from this survey will help to identify the practices of effective literacy instruction. The survey's results will be delineated by the graduate student with anonymity and confidentiality, under the supervision of Dr. Marguerite Lodico, professor from The College of Saint Rose.

Taking part in research is voluntary and you may at any time withdraw your participation. If you decide not to give permission, it will in no way influence your teaching position.

If you have any questions, please do not hesitate to contact, researcher's name or Dr. Lodico (phone number or e-mail).

I, ____________________________________, agree to participate in this study.

Date: ____________________

Appendix B

Effective ELA and Literacy Instruction Practices

You have received this survey as part of a research study. The purpose of this survey is to identify theoretical orientations methods and materials used, and assessments used in teaching literacy in those classrooms that are meeting state standards with a level of three or higher according to New York State Report Cards. Your responses are confidential and will not be shared with anyone in any way that identifies you as an individual. Only aggregated data will be presented in the final report. Your participation in this study is voluntary and will in no way influence your teaching position within your school district. Please return this survey via e-mail, no later than [insert date].

I. Demographics: Please complete the following information.

1. I teach: ____ 1st grade _____ 2nd grade _____ 3rd grade _____4th grade.
2. I have been teaching for: ____ 1–3 yrs. ____ 4–6 yrs. ____ 7–10 yrs. _____ 11+ yrs.
3. I have my ____ Bachelor's Degree ____ Master's Degree _____ Doctorate (check all that apply).
4. I have ____ 15 – 20 ____ 21–26 ____ 27 – 32 ____ 33- 38 ____39 or more students in my class.
5. Please indicate whether you teach in a self-contained class or in a collaborative setting. If collaborative please indicate specific content areas taught.

   ______ Self-Contained

   ______ Collaborative

   Content area taught ______________________________

6. Please use the following space to identify areas of expertise. Include minors or concentration areas, content study in graduate study programs, and areas of certification.

____

____

____

____

____

____

____

II. Theoretical Orientations to Teaching Literacy

Read each question below and respond by circling the number that best represents how you feel about how you use theories of teaching literacy in the classroom.

1 = Strongly Disagree;

2 = Disagree;

3 = Neutral;

4 = Agree;

5 = Strongly Agree

1. I believe that direct phonics instruction is the only instruction that should be used to teach literacy. 1 2 3 4 5
2. I believe that whole language instruction is the only instruction that should be used to teach literacy. 1 2 3 4 5
3. I believe that a blended method of literacy instruction, a program that uses phonics and whole language, should be used to teach literacy. 1 2 3 4 5

III. Materials used during literacy instruction.

Read each question below and respond by circling the number that best represents your use of materials during literacy instruction.

1 = Never;

2 = Not Usually;

3 = Sometimes;

4 = Always

1. I use a basal series for literacy instruction. 1 2 3 4
2. I use authentic texts as reading literature for literacy instruction. 1 2 3 4
3. I use authentic literature in conjunction with skills directed activities for literacy instruction. 1 2 3 4
4. I allow students to choose their own reading topics. 1 2 3 4

IV. Assessments used for identifying literacy growth.

Read each question below and respond by circling the number that best represents your use of assessments during literacy assessment.

1 = Never;

2 = Not Usually;

3 = Sometimes;

4 = Always

1. I use an end of unit skills test to assess literacy growth. 1 2 3 4
2. I use conferencing as a way of assessing growth in writing. 1 2 3 4
3. I use running records as a way of assessing fluency. 1 2 3 4

V. Comments

Please use the following space to identify any other methods, materials, or assessment tools that you use in your classroom that you believe add to your literacy instruction. You may write as much or as little as you would like and can be as specific as you would like.

___

___

___

___

___

___

___

___

___

INDEX

Page references followed by *fig* indicate an illustrated figure; followed by *t* indicate a table; followed by *b* indicate a box; followed by *e* indicate an exhibit.

D

E

J

K

L

R

S

Methods in Educational Research

Methods in Educational Research is designed to prepare students for the real world of educational research. It focuses on scientifically-based methods, school accountability, and the professional demands of the twenty-first century, empowering researchers to take an active role in conducting research in their classrooms, districts, and the greater educational community. Like the first edition, this edition helps students, educators, and researchers develop a broad and deep understanding of research methodologies. It includes substantial new content on the impact of No Child Left Behind legislation, school reform, quantitative and qualitative methodologies, logic modeling, action research, and other areas. Special features to assist the teaching and learning processes include vignettes illustrating research tied to practice, suggested readings at the end of each chapter, and discussion questions to reinforce chapter content.

Praise for the Previous Edition

"A new attempt to make this subject more relevant and appealing to students. Most striking is how useful this book is because it is really grounded in educational research. It is very well written and quite relevant for educational researchers or for the student hoping to become one."
—*PsycCRITIQUES*/American Psychological Association

"I applaud the authors for their attempt to cover a wide range of material. The straightforward language of the book helps make the material understandable for readers."
—*Journal of MultiDisciplinary Evaluation*

The Authors
Marguerite G. Lodico, EdD, is a professor at the College of Saint Rose where she teaches child development and educational research.
Dean T. Spaulding, PhD, is an associate professor at the College of Saint Rose where he teaches educational research and program evaluation.
Katherine H. Voegtle, PhD, is a professor at the College of Saint Rose where she teaches courses in human development, educational research, and educational psychology.

RESEARCH METHODS/EDUCATION

Cover design by Michael Rutkowski

ISBN 978-0-470-43680-6

90000

9 780470 436806

www.josseybass.com

JOSSEY-BASS™
An Imprint of
WILEY