

**APPLIED SOCIAL RESEARCH
METHODS SERIES**

Series Editors

LEONARD BICKMAN, Peabody College, Vanderbilt University, Nashville
DEBRA J. ROG, Vanderbilt University, Washington, DC

1. SURVEY RESEARCH METHODS
(Third Edition)
by FLOYD J. FOWLER, Jr.
2. SYNTHESIZING RESEARCH
(Third Edition)
by HARRIS COOPER
3. METHODS FOR POLICY RESEARCH
by ANN MAJCHRZAK
4. SECONDARY RESEARCH
(Second Edition)
by DAVID W. STEWART
and MICHAEL A. KAMINS
5. CASE STUDY RESEARCH
(Third Edition)
by ROBERT K. YIN
6. META-ANALYTIC PROCEDURES FOR
SOCIAL RESEARCH
(Revised Edition)
by ROBERT ROSENTHAL
7. TELEPHONE SURVEY METHODS
(Second Edition)
by PAUL J. LAVRAKAS
8. DIAGNOSING ORGANIZATIONS
(Second Edition)
by MICHAEL I. HARRISON
9. GROUP TECHNIQUES FOR
IDEA BUILDING (Second Edition)
by CARL M. MOORE
10. NEED ANALYSIS
by JACK McKILLIP
11. LINKING AUDITING AND
META EVALUATION
by THOMAS A. SCHWANDT
and EDWARD S. HALPERN
12. ETHICS AND VALUES
IN APPLIED SOCIAL RESEARCH
by ALLAN J. KIMMEL
13. ON TIME AND METHOD
by JANICE R. KELLY
and JOSEPH E. McGRATH
14. RESEARCH IN HEALTH CARE SETTINGS
by KATHLEEN E. GRADY
and BARBARA STRUDLER WALLSTON
15. PARTICIPANT OBSERVATION
by DANNY JORGENSEN
16. INTERPRETIVE INTERACTIONISM
(Second Edition)
by NORMAN K. DENZIN
17. ETHNOGRAPHY
(Second Edition)
by DAVID M. FETTERMAN
18. STANDARDIZED SURVEY INTERVIEWING
by FLOYD J. FOWLER, Jr.
and THOMAS W. MANGIONE
19. PRODUCTIVITY MEASUREMENT
by ROBERT O. BRINKERHOFF
and DENNIS E. DRESSLER
20. FOCUS GROUPS
by DAVID W. STEWART
and PREM N. SHAMDASANI
21. PRACTICAL SAMPLING
by GART T. HENRY
22. DECISION RESEARCH
by JOHN S. CARROLL
and ERIC J. JOHNSON
23. RESEARCH WITH HISPANIC
POPULATIONS
by GERARDO MARIN
and BARBARA VANOSS MARIN
24. INTERNAL EVALUATION
by ARNOLD J. LOVE
25. COMPUTER SIMULATION APPLICATIONS
by MARCIA LYNN WHICKER
and LEE SIGELMAN
26. SCALE DEVELOPMENT
by ROBERT F. DeVELLIS
27. STUDYING FAMILIES
by ANNE P. COPELAND
and KATHLEEN M. WHITE
28. EVENT HISTORY ANALYSIS
by KAZUO YAMAGUCHI
29. RESEARCH IN EDUCATIONAL SETTINGS
by GEOFFREY MARYUYAMA
and STANLEY DENO
30. RESEARCHING PERSONS WITH
MENTAL ILLNESS
by ROSALIND J. DWORKIN
31. PLANNING ETHICALLY RESPONSIBLE
RESEARCH
by JOAN E. SIEBER

**CASE
STUDY
RESEARCH**
Design and Methods

Third Edition

Robert K. Yin

Applied Social Research Methods Series
Volume 5

Other volumes in this series are listed at the back of the book

 **SAGE Publications**
International Educational and Professional Publisher
Thousand Oaks ■ London ■ New Delhi

Copyright © 2003 by Sage Publications, Inc.

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

For information:



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

Sage Publications Ltd.
6 Bonhill Street
London EC2A 4PU
United Kingdom

Sage Publications India Pvt. Ltd.
B-42 Panchsheel Enclave
Post Box 4109
New Delhi 110 017 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Yin, Robert K.

Case study research : design and methods / Robert K. Yin.- 3rd ed.

p. cm. — (Applied social research methods series ; v. 5)

Includes bibliographical references and index.

ISBN 0-7619-2552-X (alk. paper) - ISBN 0-7619-2553-8 (pbk. : alk. paper)

1. CASE method. 2. Social sciences—Research—Methodology. I. Title.

II. Series.

H62 .Y56 2002

300'.7'22—dc21

2002152696

This book is printed on acid-free paper.

03 04 05 06 10 9 8 7 6 5 4 3 2

Contents

Foreword	ix
<i>Donald T. Campbell</i>	
Preface	xiii
Additional Note About the Third Edition	xv
1. Introduction	1
The Case Study as a Research Strategy	1
Comparing Case Studies With Other Research Strategies in the Social Sciences	3
Different Kinds of Case Studies, but a Common Definition	11
Summary	15
2. Designing Case Studies	19
General Approach to Designing Case Studies	19
Criteria for Judging the Quality of Research Designs	33
Case Study Designs	39
Modest Advice in Selecting Case Study Designs	53
3. Conducting Case Studies: Preparing for Data Collection	57
The Case Study Investigator: Desired Skills	58
Training and Preparation for a Specific Case Study	62
The Case Study Protocol	67
Screening Case Study Nominations	77
The Pilot Case Study	78
Summary	81
4. Conducting Case Studies: Collecting the Evidence	83
Six Sources of Evidence	85
Three Principles of Data Collection	97
Summary	106
5. Analyzing Case Study Evidence	109
An Analytic Strategy: More Than Familiarity With Analytic Tools	109
Specific Analytic Techniques	116

Editorial Assistant: Karen Wiley
Production Editor: Sanford Robinson
Copy Editor: Gillian Dickens
Typesetter: C&M Digitals (P) Ltd.
Indexer: Molly Hall

NOTES

1. The discussion only pertains to the use of these strategies in the social sciences, making no claims for commenting on the use of experiments, for instance, in physics, astronomy, or other fields.

2. Additional examples of explanatory case studies are presented in their entirety in a companion book, *Applications of Case Study Research* (Yin, 2003), in Chapters 4, 5, 6, and 7. Similarly, two examples of descriptive case studies are presented in their entirety in Chapters 2 and 3 of the same book.

3. Stake (1994) has a similar approach to defining case studies. He considers them not to be “a methodological choice but a choice of object to be studied.” Furthermore, the object must be a “functioning specific” (such as a person or classroom) but not a generality (such as a policy).

2

Designing Case Studies

A research design is the logic that links the data to be collected (and the conclusions to be drawn) to the initial questions of study. Every empirical study has an implicit, if not explicit, research design. Articulating “theory” about what is being studied helps to operationalize case study designs and make them more explicit.

In addition, the development of case study designs needs to maximize four conditions related to design quality: (a) construct validity, (b) internal validity (for explanatory or causal case studies only), (c) external validity, and (d) reliability. How investigators deal with these aspects of quality control is summarized in Chapter 2 but also is a major theme throughout the remainder of the book.

Among the actual case study designs, four major types are relevant, following a 2×2 matrix. The first pair consists of single-case and multiple-case designs. The second pair, which can occur in combination with either of the first pair, is based on the unit or units of analysis to be covered—and distinguishes between holistic and embedded designs. Among these designs, most multiple-case designs are likely to be stronger than single-case designs. Trying to use even a “two-case” design is therefore a worthy objective compared to doing a single-case study.

**GENERAL APPROACH TO
DESIGNING CASE STUDIES**

In identifying the research strategy for your research project, Chapter 1 has shown when you should select the case study strategy, as opposed to other strategies. The next task is to design your case study. For this purpose, as in designing any other type of research investigation, a plan, or *research design*, is needed.

The development of this research design is a difficult part of doing case studies. Unlike other research strategies, a comprehensive “catalog” of research designs for case studies has yet to be developed. There are no textbooks, like those in the biological and psychological sciences, covering

such design considerations as the assignment of subjects to different “groups,” the selection of different stimuli or experimental conditions, or the identification of various response measures (see Cochran & Cox, 1957; Fisher, 1935, cited in Cochran & Cox, 1957; Sidowski, 1966). In a laboratory experiment, each of these choices reflects an important logical connection to the issues being studied. Similarly, there are not even textbooks like the well-known volumes by Campbell and Stanley (1966) or by Cook and Campbell (1979) that summarize the various research designs for quasi-experimental situations. Nor have there emerged any common designs—for example, “panel” studies—such as those now recognized in doing survey research (see Kidder & Judd, 1986, chap. 6).

One pitfall to be avoided, however, is to consider case study designs to be a subset or variant of the research designs used for other strategies, such as experiments. For the longest time, scholars incorrectly thought that the case study was but one type of quasi-experimental design (the one-shot, posttest-only design). This misperception has finally been corrected, with the following statement appearing in a revision on quasi-experimental designs: “Certainly the case study as normally practiced should not be demeaned by identification with the one-group post-test-only design” (Cook & Campbell, 1979, p. 96). In other words, the one-shot, posttest-only design as a quasi-experimental design still may be considered flawed, but the case study has now been recognized as something different. In fact, the case study is a separate research method that has its own research designs.

Unfortunately, case study research designs have not been codified. The following chapter therefore expands on the new methodological ground broken by earlier editions of this book and describes a basic set of research designs for doing single- and multiple-case studies. Although these designs will need to be continually modified and improved in the future, in their present form, they will nevertheless help you to design more rigorous and methodologically sound case studies.

Definition of Research Designs

Every type of empirical research has an implicit, if not explicit, research design. In the most elementary sense, the design is the logical sequence that connects the empirical data to a study’s initial research questions and, ultimately, to its conclusions. Colloquially, a research design is a *logical plan for getting from here to there*, where *here* may be defined as the initial set of questions to be answered, and *there* is some set of conclusions (answers) about these questions. Between “here” and “there” may be found a number of major steps, including the collection and analysis of relevant data. As a

summary definition, another textbook has described a research design as a plan that

guides the investigator in the process of collecting, analyzing, and interpreting observations. It is a *logical model of proof* that allows the researcher to draw inferences concerning causal relations among the variables under investigation. (Nachmias & Nachmias, 1992, pp. 77-78, emphasis added)

Another way of thinking about a research design is as a “blueprint” of research, dealing with at least four problems: what questions to study, what data are relevant, what data to collect, and how to analyze the results (Philliber, Schwab, & Samsloss, 1980).

Note that a research design is much more than a work plan. The main purpose of the design is to help to avoid the situation in which the evidence does not address the initial research questions. In this sense, a research design deals with a *logical* problem and not a *logistical* problem. As a simple example, suppose you want to study a single organization. Your research questions, however, have to do with the organization’s relationships with other organizations—their competitive or collaborative nature, for example. Such questions can be answered only if you collect information directly from the other organizations and not merely from the one you started with. If you complete your study by examining only one organization, you cannot draw accurate conclusions about interorganizational partnerships. This is a flaw in your research design, not in your work plan. The outcome could have been avoided if you had developed an appropriate research design in the first place.

Components of Research Designs

For case studies, five components of a research design are especially important:

1. a study’s questions;
2. its propositions, if any;
3. its unit(s) of analysis;
4. the logic linking the data to the propositions; and
5. the criteria for interpreting the findings.

Study questions. This first component has already been described in Chapter 1. Although the substance of your questions will vary, Chapter 1 suggested that the *form* of the question—in terms of “who,” “what,”

“where,” “how,” and “why”—provides an important clue regarding the most relevant research strategy to be used. The case study strategy is most likely to be appropriate for “how” and “why” questions, so your initial task is to clarify precisely the nature of your study questions in this regard.

Study propositions. As for the second component, each proposition directs attention to something that should be examined within the scope of study. For instance, assume that your research, on the topic of interorganizational partnerships, began with the following question: How and why do organizations collaborate with one another to provide joint services (for example, a manufacturer and a retail store collaborating to sell certain computer products)? These “how” and “why” questions, capturing what you are really interested in answering, led you to the case study as the appropriate strategy in the first place. Nevertheless, these “how” and “why” questions do not point to what you should study.

Only if you are forced to state some propositions will you move in the right direction. For instance, you might think that organizations collaborate because they derive mutual benefits. This proposition, besides reflecting an important theoretical issue (that other incentives for collaboration do not exist or are unimportant), also begins to tell you where to look for relevant evidence (to define and ascertain the extent of specific benefits to each organization).

At the same time, some studies may have a legitimate reason for not having any propositions. This is the condition—which exists in experiments, surveys, and the other research strategies alike—in which a topic is the subject of “exploration.” Every exploration, however, should still have some purpose. Instead of propositions, the design for an exploratory study should state this purpose, as well as the criteria by which an exploration will be judged successful. Consider the analogy in BOX 4 for exploratory case studies. Can you imagine how you would ask for support from Queen Isabella to do your exploratory study?

Unit of analysis. This third component is related to the fundamental problem of defining what the “case” is—a problem that has plagued many investigators at the outset of case studies. For instance, in the classic case study, a “case” may be an individual. Jennifer Platt (1992a, 1992b) has noted how the early case studies in the Chicago school of sociology were life histories of such roles as juvenile delinquents or derelict men. You also can imagine case studies of clinical patients, exemplary students, or political leaders. In each situation, an individual person is the case being studied, and the individual is the primary unit of analysis. Information about each

BOX 4

“Exploration” as an Analogy for an Exploratory Case Study

When Christopher Columbus went to Queen Isabella to ask for support for his “exploration” of the New World, he had to have some reasons for asking for three ships (Why not one? Why not five?), and he had to have some rationale for going westward (Why not south? Why not south and then east?). He also had some (mistaken) criteria for recognizing the Indies when he actually encountered it. In short, his exploration began with some rationale and direction, even if his initial assumptions might later have been proved wrong (Wilford, 1992). This same degree of rationale and direction should underlie even an exploratory case study.

relevant individual would be collected, and several such individuals or “cases” might be included in a multiple-case study. Propositions would still be needed to help identify the relevant information about this individual or individuals. Without such propositions, an investigator might be tempted to cover “everything,” which is impossible to do. For example, the propositions in studying these individuals might involve the influence of early childhood or the role of peer relationships. Such topics already represent a vast narrowing of the relevant data. The more a study contains specific propositions, the more it will stay within feasible limits.

Of course, the “case” also can be some event or entity that is less well defined than a single individual. Case studies have been done about decisions, programs, the implementation process, and organizational change. Feagin et al. (1991) contains some classic examples of these single cases in sociology and political science. Beware of these types of topics—none is easily defined in terms of the beginning or end points of the “case.” For example, a case study of a specific program may reveal (a) variations in program definition, depending on the perspective of different actors, and (b) program components that preexisted the formal designation of the program. Any case study of such a program would therefore have to confront these conditions in delineating the unit of analysis.

As a general guide, your tentative definition of the unit of analysis (and therefore of the case) is related to the way you have defined your initial research questions. Suppose, for example, you want to study the role of the United States in the world economy. Peter Drucker (1986) has written a provocative essay about fundamental changes in the world economy, including the importance of “capital movements” independent of the flow

of goods and services. The unit of analysis for your case study might be a country's economy, an industry in the world marketplace, an economic policy, or the trade or capital flow between two countries. Each unit of analysis would call for a slightly different research design and data collection strategy.

Selection of the appropriate unit of analysis will occur when you accurately specify your primary research questions. If your questions do not lead to the favoring of one unit of analysis over another, your questions are probably either too vague or too numerous—and you may have trouble conducting your case study. However, when you have arrived at a definition of the unit of analysis, do not consider closure permanent. Your choice of the unit of analysis, as with other facets of your research design, can be revisited as a result of discoveries arising during your data collection (see discussion and cautions about flexibility, throughout and at the end of this chapter).

Sometimes, the unit of analysis may have been defined one way, even though the phenomenon being studied calls for a different definition. Most frequently, investigators have confused case studies of neighborhoods with case studies of small groups (as another example, confusing a new technology with the workings of an engineering team in an organization; see BOX 5A). How a geographic *area* such as a neighborhood copes with racial transition, upgrading, and other phenomena can be quite different from how a small *group* copes with these same phenomena. For instance, *Street Corner Society* (Whyte, 1943/1955—also see BOX 2 in Chapter 1 of this book) and *Tally's Corner* (Liebow, 1967—also see BOX 9, this chapter) have often been mistaken for being case studies of urban neighborhoods when in fact they are case studies of small groups (note that in neither book is the neighborhood geography described, even though the small groups lived in a small area with clear neighborhood implications). BOX 5B, however, presents a good example of how units of analyses can be defined in a more discriminating manner—in the field of world trade.

Most investigators will encounter this type of confusion in defining the unit of analysis. To reduce the confusion, one recommended practice is to discuss the potential case with a colleague. Try to explain to that person what questions you are trying to answer and why you have chosen a specific case or group of cases as a way of answering those questions. This may help you to avoid incorrectly identifying the unit of analysis.

Once the general definition of the case has been established, other clarifications in the unit of analysis become important. If the unit of analysis is a small group, for instance, the persons to be included within the group (the immediate topic of the case study) must be distinguished from those who are outside it (the context for the case study). Similarly, if the case is about

BOX 5A

What Is the Unit of Analysis?

The Soul of a New Machine was a Pulitzer Prize-winning book by Tracy Kidder (1981). The book, also a best-seller, is about the development of a new computer produced by Data General Corporation, intended to compete directly with one produced by Digital Equipment Corporation.

This easy-to-read book describes how Data General's engineering team invented and developed the new computer. The book begins with the initial conceptualization of the computer and ends when the engineering team relinquished control of the machine to Data General's marketing staff.

The book is an excellent example of a case study. However, the book also illustrates a fundamental problem in doing case studies—that of defining the *unit of analysis*. Is the case study about the computer, or is it about the dynamics of a small group—the engineering team? The answer is critical if we want to understand how the case study relates to a broader body of knowledge—that is, whether to generalize to a technology topic or to a group dynamics topic. Because the book is not an academic study, it does not need to, nor does it, provide an answer.

BOX 5B

A Clearer Choice Among Units of Analysis

Ira Magaziner and Mark Patinkin's (1989) book, *The Silent War: Inside the Global Business Battles Shaping America's Future*, presents nine case studies. Each case study helps the reader to understand a real-life situation of international economic competition.

Two of the cases appear similar but in fact have different main units of analysis. One case, about the Korean firm Samsung, is a case study of the critical policies that make the firm competitive. Understanding Korean economic development is part of the context, and the case study also contains an embedded unit—Samsung's development of the microwave oven as an illustrative product. The other case, about the development of an Apple computer factory in Singapore, is in fact a case study of Singapore's critical policies that make the country competitive. The Apple computer factory experience—an embedded unit of analysis—is actually an illustrative example of how the national policies affect foreign investments.

These two cases show how the definition of the main and embedded units of analyses, as well as the definition of the contextual events surrounding these units, depends on the level of inquiry. The main unit of analysis is likely to be at the level being addressed by the main study questions.

local services in a specific geographic area, decisions need to be made about those services whose district boundaries do not coincide with the area. Finally, for almost any topic that might be chosen, specific time boundaries are needed to define the beginning and end of the case. All of these types of questions need to be considered and answered to define the unit of analysis and thereby to determine the limits of the data collection and analysis.

One final point, pertaining to the role of the available research literature, needs to be made about defining the case and the unit of analysis. Most researchers will want to compare their findings with previous research; for this reason, key definitions used in your study should not be idiosyncratic. Rather, each case study and unit of analysis either should be similar to those previously studied by others or should innovate in clear, operationally defined ways. In this manner, the previous literature also can become a guide for defining the case and unit of analysis.

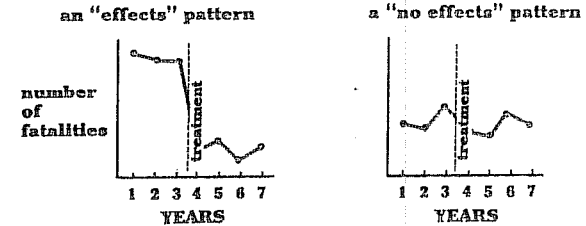
Linking data to propositions and criteria for interpreting the findings. The fourth and fifth components have been the least well developed in case studies. These components foreshadow the data analysis steps in case study research, and a research design should lay a solid foundation for this analysis.

Linking data to propositions can be done any number of ways, but none has become as precisely defined as the assignment of subjects and treatment conditions in psychological experiments (which is one way that hypotheses and data are connected in psychology). One promising approach for case studies is the idea of “pattern matching” described by Donald Campbell (1975), whereby several pieces of information from the same case may be related to some theoretical proposition (also see Chapter 5 of this book).

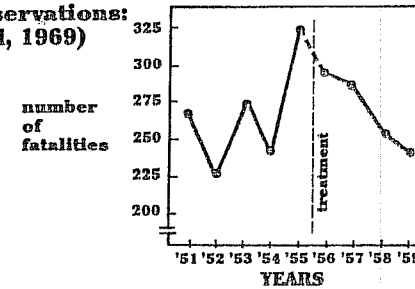
In a related article on one type of pattern—a time-series pattern—Campbell (1969) illustrates this approach. In the article, Campbell first showed how the annual number of traffic fatalities in Connecticut had seemed to decline after the passage of a new state law limiting the speed to 55 miles per hour. However, further examination of the fatality rate, over a number of years before and after the legal change, showed unsystematic fluctuation rather than any marked reduction. A simple eyeball test was all that was needed to show that the actual pattern *looked* unsystematic rather than following a downtrend (see Figure 2.1), and thus Campbell concluded that the speed limit had had no effect on the number of traffic fatalities.

What Campbell did was to describe two potential patterns and then show that the data matched one better than the other. If the two potential patterns are considered rival propositions (an “effects” proposition and a “no

a priori propositions:



actual observations: (Campbell, 1969)



... a “no effects” pattern

Figure 2.1 An Example of Pattern Matching
SOURCE: COSMOS Corporation.

effects” proposition, regarding the impact of the new speed limit law), the pattern-matching technique is a way of relating the data to the propositions, even though the entire study consists of only a single case (the state of Connecticut).

This article also illustrates the problems in dealing with the fifth component, *the criteria for interpreting a study's findings*. Campbell's (1969) data matched one pattern better than they matched the other. But how close does a match have to be in order to be considered a match? Note that Campbell did not do any statistical test to make the comparison. Nor would a statistical test have been possible because each data point in the pattern was a single number—the number of fatalities for that year—for which one could not calculate a variance or conduct any statistical test. Currently, there is no precise way of setting the criteria for interpreting these types of findings. One hopes that the different patterns are sufficiently contrasting (as in Campbell's case) that the findings can be interpreted in terms of comparing at least two rival propositions. (Much more about the importance of such rivals appears in Chapter 5.)

Summary. A research design should include five components. Although the current state of the art does not provide detailed guidance on the last two, the complete research design should not only indicate what data are to be collected—as indicated by (a) a study’s questions, (b) its propositions, and (c) its units of analysis. The design also should tell you what is to be done after the data have been collected—as indicated by (d) the logic linking the data to the propositions and (e) the criteria for interpreting the findings.

The Role of Theory in Design Work

Covering these preceding five components of research designs will effectively force you to begin constructing a preliminary theory related to your topic of study. This role of theory development, prior to the conduct of any data collection, is one point of difference between case studies and related methods such as ethnography (Lincoln & Guba, 1985, 1986; Van Maanen, 1988; Van Maanen et al., 1982) and “grounded theory” (Strauss & Corbin, 1998). Typically, these related methods deliberately avoid specifying any theoretical propositions at the outset of an inquiry. As a result, students confusing these methods with case studies wrongly think that by having selected the case study method, they can proceed quickly into the data collection phase of their work, and they may have been encouraged to make their “field contacts” as quickly as possible. No guidance could be more misleading. Among other considerations, the relevant field contacts depend on an understanding—or theory—of what is being studied.

Theory development. For case studies, theory development as part of the design phase is essential, whether the ensuing case study’s purpose is to develop or test theory. Using a case study on the implementation of a new management information system (MIS) as an example (Markus, 1983), the simplest ingredient of a theory is a statement such as the following:

The case study will show why implementation only succeeded when the organization was able to re-structure itself, and not just overlay the new MIS on the old organizational structure. (Markus, 1983)

The statement presents the nutshell of a theory of MIS implementation—that is, that organizational restructuring is needed to make MIS implementation work.

Using the same case, an additional ingredient might be the following statement:

The case study will also show why the simple replacement of key persons was not sufficient for successful implementation. (Markus, 1983)

This second statement presents the nutshell of a *rival* theory—that is, that MIS implementation fails because of the resistance to change on the part of individual people and that the replacement of such people is the only requirement for implementation to succeed.

You can see that as these two initial ingredients are elaborated, the stated ideas will increasingly cover the questions, propositions, units of analysis, logic connecting data to propositions, and criteria for interpreting the findings—that is, the five components of the needed research design. In this sense, the complete research design embodies a “theory” of what is being studied. This theory should by no means be considered with the formality of grand theory in social science, nor are you being asked to be a masterful theoretician. Rather, the simple goal is to have a sufficient blueprint for your study, and this requires theoretical propositions, usefully noted by two authors as “a [hypothetical] story about why acts, events, structure, and thoughts occur” (Sutton & Staw, 1995, p. 378). Then, the complete research design will provide surprisingly strong guidance in determining what data to collect and the strategies for analyzing the data. For this reason, theory development prior to the collection of any case study data is an essential step in doing case studies.

However, theory development takes time and can be difficult (Eisenhardt, 1989). For some topics, existing works may provide a rich theoretical framework for designing a specific case study. If you are interested in international economic development, for instance, Peter Drucker’s (1986) “The Changed World Economy” is an exceptional source of theories and hypotheses. Drucker claims that the world economy had changed significantly from the past. He points to the “uncoupling” between the primary products (raw materials) economy and the industrial economy, a similar uncoupling between low labor costs and manufacturing production, and the uncoupling between financial markets and the real economy of goods and services. To test these propositions might require different studies, some focusing on the different uncouplings, others focusing on specific industries, and yet others explaining the plight of specific countries. Each different study would likely call for a different unit of analysis. Drucker’s theoretical framework would provide guidance for designing these studies and even for collecting relevant data.

In other situations, the appropriate theory may be a descriptive theory (see BOX 6 and also BOX 2 earlier for another example), and your concern should focus on such issues as (a) the purpose of the descriptive effort,

BOX 6

Using a Metaphor to Develop Descriptive Theory

Whether four “countries”—the American colonies, Russia, England, and France—all underwent similar courses of events during their major political revolutions is the topic of Crane Brinton’s (1938) famous historical study, *The Anatomy of a Revolution*. Tracing and analyzing these events is done in a descriptive manner, as the author’s purpose is not so much to explain the revolutions as to determine whether they followed similar courses.

The “cross-case” analysis reveals major similarities: All societies were on the upgrade (not downgrade, as might have been expected) economically; there were bitter class antagonisms; the intellectuals deserted their governments; government machinery was inefficient; and the ruling class exhibited immoral, dissolute, or inept behavior (or all three). However, rather than relying solely on this “factors” approach to description, the author also develops the metaphor of a human body suffering from a fever as a way of describing the pattern of events over time. The author adeptly uses the cyclic pattern of fever and chills, rising to a critical point and followed by a false tranquility, to describe the ebb and flow of events in the four revolutions.

(b) the full but realistic range of topics that might be considered a “complete” description of what is to be studied, and (c) the likely topic(s) that will be the essence of the description. Good responses to these issues, including the rationales underlying the responses, will help you go a long way toward developing the needed theoretical base—and research design—for your study.

For yet other topics, the existing knowledge base may be poor, and the available literature will provide no conceptual framework or hypotheses of note. Such a knowledge base does not lend itself to the development of good theoretical statements, and any new empirical study is likely to assume the characteristic of an “exploratory” study. Nevertheless, as noted earlier with the illustrative case in BOX 4, even an exploratory case study should be preceded by statements about (a) what is to be explored, (b) the purpose of the exploration, and (c) the criteria by which the exploration will be judged successful.¹

Illustrative types of theories. In general, to overcome the barriers to theory development, you should try to prepare for your case study by doing such things as reviewing the literature related to what you would like to

study (also see Cooper, 1984), discussing your topic and ideas with colleagues or teachers, and asking yourself challenging questions about what you are studying, why you are proposing to do the study, and what you hope to learn as a result of the study.

As a further reminder, you should be aware of the full range of theories that might be relevant to your study. For instance, note that the MIS example illustrates MIS “implementation” theory, which is but one type of theory that can be the subject of study. Other types of theories for you to consider include the following:

- *Individual theories*—for example, theories of individual development, cognitive behavior, personality, learning and disability, individual perception, and interpersonal interactions
- *Group theories*—for example, theories of family functioning, informal groups, work teams, supervisory-employee relations, and interpersonal networks
- *Organizational theories*—for example, theories of bureaucracies, organizational structure and functions, excellence in organizational performance, and interorganizational partnerships
- *Societal theories*—for example, theories of urban development, international behavior, cultural institutions, technological development, and marketplace functions

Other examples cut across some of these illustrative types. Decision-making theory (Carroll & Johnson, 1992), for instance, can involve individuals, organizations, or social groups. As another example, a common topic of case studies is the evaluation of publicly supported programs, such as federal, state, or local programs. In this situation, the development of a theory of how a program is supposed to work is essential to the design of the evaluation but has been commonly underemphasized in the past (Bickman, 1987). According to Bickman (1987), analysts have frequently confused the theory of the program (e.g., how to make education more effective) with the theory of program implementation (e.g., how to install an effective program). Where policymakers want to know the desired substantive steps (e.g., describe a newly effective curriculum), the analysts unfortunately recommend managerial steps (e.g., hire a good project director). This mismatch can be avoided by giving closer attention to the substantive theory.

Generalizing from case study to theory. Theory development does not only facilitate the data collection phase of the ensuing case study. The appropriately developed theory also is the level at which the generalization of the case study results will occur. This role of theory has been characterized

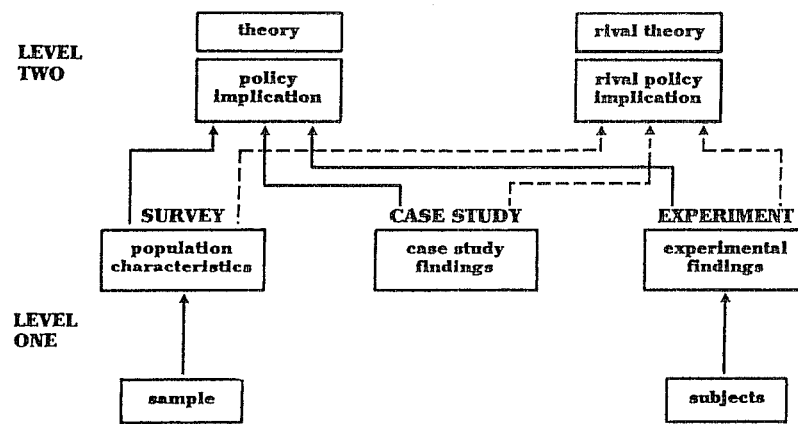


Figure 2.2 Making Inferences: Two Levels
SOURCE: COSMOS Corporation.

throughout this book as “analytic generalization” and has been contrasted with another way of generalizing results, known as “statistical generalization.” Understanding the distinction between these two types of generalization may be your most important challenge in doing case studies.

Let us first take the more commonly recognized way of generalizing—“statistical generalization”—although it is the less relevant one for doing case studies. In statistical generalization, an inference is made about a population (or universe) on the basis of empirical data collected about a sample. This is shown as a *Level One Inference* in Figure 2.2.² This method of generalizing is commonly recognized because research investigators have ready access to quantitative formulas for determining the confidence with which generalizations can be made, depending mostly on the size and internal variation within the universe and sample. Moreover, this is the most common way of generalizing when doing surveys (e.g., Fowler, 1988; Lavrakas, 1987) or analyzing archival data.

A fatal flaw in doing case studies is to conceive of statistical generalization as the method of generalizing the results of the case study. This is because your cases are not “sampling units” and should not be chosen for this reason. Rather, individual case studies are to be selected as a laboratory investigator selects the topic of a new experiment. Multiple cases, in this sense, should be considered like multiple experiments. Under these circumstances, the mode of generalization is “analytic generalization,” in which a previously developed theory is used as a template with which to compare the

empirical results of the case study. If two or more cases are shown to support the same theory, replication may be claimed. The empirical results may be considered yet more potent if two or more cases support the same theory but do not support an equally plausible, *rival* theory. Graphically, this type of generalization is shown as a *Level Two Inference* in Figure 2.2.

Analytic generalization can be used whether your case study involves one or several cases, which shall be later referenced as single-case or multiple-case studies. Furthermore, the logic of replication and the distinction between statistical and analytic generalization will be covered in greater detail in the discussion of multiple-case study designs. The main point at this juncture is that you should try to aim toward analytic generalization in doing case studies, and you should avoid thinking in such confusing terms as “the sample of cases” or the “small sample size of cases,” as if a single case study were like a single respondent in a survey or a single subject in an experiment. In other words, in terms of Figure 2.2, you should aim for *Level Two Inferences* when doing case studies.

Because of the importance of this distinction between the two ways of generalizing, you will find repeated examples and discussion throughout the remainder of this chapter as well as in Chapter 5.

Summary: This subsection has suggested that a complete research design, covering the five components described earlier, in fact benefits from the development of a theoretical framework for the case study that is to be conducted. Rather than resisting such a requirement, a good case study investigator should make the effort to develop this theoretical framework, no matter whether the study is to be explanatory, descriptive, or exploratory. The use of theory, in doing case studies, is not only an immense aid in defining the appropriate research design and data collection but also becomes the main vehicle for generalizing the results of the case study.

CRITERIA FOR JUDGING THE QUALITY OF RESEARCH DESIGNS

Because a research design is supposed to represent a logical set of statements, you also can judge the quality of any given design according to certain logical tests. Concepts that have been offered for these tests include trustworthiness, credibility, confirmability, and data dependability (U.S. General Accounting Office, 1990).

Four tests, however, have been commonly used to establish the quality of any empirical social research. Because case studies are one form of such

Tests	Case Study Tactic	Phase of research in which tactic occurs
Construct validity	● Use multiple sources of evidence	data collection
	● Establish chain of evidence	data collection
	● Have key informants review draft case study report	composition
Internal validity	● Do pattern-matching	data analysis
	● Do explanation-building	data analysis
	● Address rival explanations	data analysis
	● Use logic models	data analysis
External validity	● Use theory in single-case studies	research design
	● Use replication logic in multiple-case studies	research design
Reliability	● Use case study protocol	data collection
	● Develop case study database	data collection

Figure 2.3 Case Study Tactics for Four Design Tests
SOURCE: COSMOS Corporation.

research, the four tests also are relevant to case studies. An important innovation of this book is the identification of several tactics for dealing with these four tests when doing case studies. Figure 2.3 lists the four widely used tests and the recommended case study tactics, as well as a cross-reference to the phase of research when the tactic is to be used. (Each tactic is described in detail in the relevant chapter of this book.)

Because the four tests are common to all social science methods, the tests have been summarized in numerous textbooks (see Kidder & Judd, 1986, pp. 26-29):

- *Construct validity*: establishing correct operational measures for the concepts being studied
- *Internal validity* (for explanatory or causal studies only, and not for descriptive or exploratory studies): establishing a causal relationship, whereby certain conditions are shown to lead to other conditions, as distinguished from spurious relationships
- *External validity*: establishing the domain to which a study's findings can be generalized
- *Reliability*: demonstrating that the operations of a study—such as the data collection procedures—can be repeated, with the same results

This list is more complex than the standard “validity” and “reliability” concepts to which most students have been exposed, and each item deserves explicit attention. For case studies, an important revelation is that the several tactics to be used in dealing with these tests should be applied throughout the subsequent conduct of the case study and not just at the beginning. In this sense, “design work” actually continues beyond the initial design plans.

Construct Validity

This first test is especially problematic in case study research. People who have been critical of case studies often point to the fact that a case study investigator fails to develop a sufficiently operational set of measures and that “subjective” judgments are used to collect the data.³ Take an example such as studying “neighborhood change”—a common case study topic (e.g., Bradshaw, 1999; Keating & Krumholz, 1999).

Over the years, concerns have arisen over how certain urban neighborhoods have changed their character. Any number of case studies have examined the types of changes and their consequences. However, without prior specification of the significant, operational events that constitute “change,” a reader cannot tell whether the recorded changes in a case study genuinely reflect critical events in a neighborhood or whether they happen to be based on an investigator's impressions only.

Neighborhood change can indeed cover a wide variety of phenomena: racial turnover, housing deterioration and abandonment, changes in the pattern of urban services, shifts in a neighborhood's economic institutions, or the turnover from low- to middle-income residents in revitalizing neighborhoods. To meet the test of construct validity, an investigator must be sure to cover two steps:

1. Select the specific types of changes that are to be studied (and relate them to the original objectives of the study) and
2. Demonstrate that the selected measures of these changes do indeed reflect the specific types of change that have been selected.

For example, suppose you satisfy the first step by stating that you plan to study the rise in neighborhood crime. The second step now demands that you also justify why you might be using police-reported crime (which happens to be the standard measure used in the FBI Uniform Crime Reports) as your measure of crime. Perhaps this is not a valid measure, given that large proportions of crimes are not reported to the police.

As Figure 2.3 shows, three tactics are available to increase construct validity when doing case studies. The first is the use of *multiple sources of evidence*, in a manner encouraging convergent lines of inquiry, and this tactic is relevant during data collection (see Chapter 4). A second tactic is to establish a *chain of evidence*, also relevant during data collection (see also Chapter 4). The third tactic is to have the draft case study report reviewed by key informants (a procedure described further in Chapter 6).

Internal Validity

This second test has been given the greatest attention in experimental and quasi-experimental research (see Campbell & Stanley, 1966; Cook & Campbell, 1979). Numerous “threats” to validity have been identified, mainly dealing with spurious effects. However, because so many textbooks already cover this topic, only two points need to be made here.

First, internal validity is only a concern for causal (or explanatory) case studies, in which an investigator is trying to determine whether event x led to event y . If the investigator incorrectly concludes that there is a causal relationship between x and y without knowing that some third factor— z —may actually have caused y , the research design has failed to deal with some threat to internal validity. Note that this logic is inapplicable to descriptive or exploratory studies (whether the studies are case studies, surveys, or experiments), which are not concerned with making causal claims.

Second, the concern over internal validity, for case study research, may be extended to the broader problem of making inferences. Basically, a case study involves an inference every time an event cannot be directly observed. An investigator will “infer” that a particular event resulted from some earlier occurrence, based on interview and documentary evidence collected as part of the case study. Is the inference correct? Have all the rival explanations and possibilities been considered? Is the evidence convergent? Does it appear to be airtight? A research design that has anticipated these questions has begun to deal with the overall problem of making inferences and therefore the specific problem of internal validity.

However, the specific tactics for achieving this result are difficult to identify. This is especially true in doing case studies. As one set of suggestions, Figure 2.3 shows that the analytic tactic of *pattern matching*, already touched on but to be described further in Chapter 5, is one way of addressing internal validity. Three other analytic tactics—*explanation building*, *addressing rival explanations*, and *using logic models*—also are described in Chapter 5.

External Validity

The third test deals with the problem of knowing whether a study’s findings are generalizable beyond the immediate case study. In the simplest example, if a study of neighborhood change focused on one neighborhood, are the results applicable to another neighborhood? The external validity problem has been a major barrier in doing case studies. Critics typically state that single cases offer a poor basis for generalizing. However, such critics are implicitly contrasting the situation to survey research, in which a sample (if selected correctly) readily generalizes to a larger universe. *This analogy to samples and universes is incorrect when dealing with case studies.* Survey research relies on *statistical* generalization, whereas case studies (as with experiments) rely on *analytical* generalization. In analytical generalization, the investigator is striving to generalize a particular set of results to some broader theory (see BOX 7).

For example, the theory of neighborhood change that led to a case study in the first place is the same theory that will help to identify the other cases to which the results are generalizable. If a study had focused on population transition in an urban neighborhood (e.g., Flippen, 2001), the procedure for selecting a neighborhood for study would have begun with identifying those types of neighborhoods within which transitions were occurring. Theories about population transition would then be the domain to which the results could later be generalized.

The generalization is not automatic, however. A theory must be tested by replicating the findings in a second or even a third neighborhood, where the theory has specified that the same results should occur. Once such direct replications have been made, the results might be accepted as providing strong support for the theory, even though further replications had not been performed. This *replication logic* is the same that underlies the use of experiments (and allows scientists to cumulate knowledge across experiments) and, as shown in Figure 2.3, will be discussed further in this chapter in the section on multiple-case designs.

Reliability

Most people are probably already familiar with this final test. The objective is to be sure that if a later investigator followed the same procedures as described by an earlier investigator and conducted the same case study all over again, the later investigator should arrive at the same findings and conclusions. (Note that the emphasis is on doing the *same* case over again, not on “replicating” the results of one case by doing another case study.) The goal of reliability is to minimize the errors and biases in a study.

BOX 7

How Case Studies Can Be Generalized to Theory

A common complaint about case studies is that it is difficult to generalize from one case to another. Thus, analysts fall into the trap of trying to select a “representative” case or set of cases. Yet no set of cases, no matter how large, is likely to deal satisfactorily with the complaint.

The problem lies in the very notion of generalizing to other case studies. Instead, an analyst should try to generalize findings to “theory,” analogous to the way a scientist generalizes from experimental results to theory. (Note that the scientist does not attempt to select “representative” experiments.)

This approach is well illustrated by Jane Jacobs (1961) in her famous book, *The Death and Life of Great American Cities*. The book is based mostly on experiences from New York City. However, the chapter topics, rather than reflecting the single experiences of New York, cover broader theoretical issues in urban planning, such as the role of sidewalks, the role of neighborhood parks, the need for primary mixed uses, the need for small blocks, and the processes of slumming and unslumming. In the aggregate, these issues in fact represent the building of a theory of urban planning.

Jacobs’s book created heated controversy in the planning profession. As a partial result, new empirical inquiries were made in other locales to examine one or another facet of her rich and provocative ideas. Her *theory*, in essence, became the vehicle for examining other cases, and the theory still stands as a significant contribution to the field of urban planning.

One prerequisite for allowing this other investigator to repeat an earlier case study is to document the procedures followed in the earlier case. Without such documentation, you could not even repeat your own work (which is another way of dealing with reliability). In the past, case study research procedures have been poorly documented, making external reviewers suspicious of the reliability of the case study.⁴ As specific tactics to overcome these shortcomings, Chapter 3 discusses the use of a *case study protocol* to deal with the documentation problem in detail, and Chapter 4 describes another tactic, the development of a *case study database* (see Figure 2.3).

The general way of approaching the reliability problem is to make as many steps as operational as possible and to conduct research as if someone were always looking over your shoulder. In accounting and bookkeeping, one is always aware that any calculations must be capable of being

audited. In this sense, an auditor is also performing a reliability check and must be able to produce the same results if the same procedures are followed. A good guideline for doing case studies is therefore to conduct the research so that an auditor could repeat the procedures and arrive at the same results.

Summary. Four tests may be considered relevant in judging the quality of a research design. In designing and doing case studies, various tactics are available to deal with these tests, though not all of the tactics occur at the formal stage of designing a case study. Some of the tactics occur during the data collection, data analysis, or compositional phases of the research and are therefore described in greater detail in subsequent chapters of this book.

CASE STUDY DESIGNS

These general characteristics of research designs serve as a background for considering the specific designs for case studies. Four types of designs will be discussed, based on a 2×2 matrix (see Figure 2.4). The matrix first shows that every type of design will include the desire to analyze contextual conditions in relation to the “case,” and the dotted lines between the two indicate that the boundaries between the case and the context are not likely to be sharp. The matrix then shows that single- and multiple-case studies reflect different design situations and that within these two variants, there also can be a unitary unit or multiple units of analysis. The resulting four types of designs for case studies are single-case (holistic) designs (Type 1), single-case (embedded) designs (Type 2), multiple-case (holistic) designs (Type 3), and multiple-case (embedded) designs (Type 4). The rationale for these four types of designs is as follows.

What Are the Potential Single-Case Designs (Types 1 and 2)?

Rationale for single-case designs. A primary distinction in designing case studies is between *single-* and *multiple-*case designs. This means the need for a decision, prior to any data collection, on whether a single case study or multiple cases are going to be used to address the research questions. The single-case study is an appropriate design under several circumstances, and five rationales are given below. Recall that a single-case study is analogous to a single experiment, and many of the same conditions that justify a single experiment also justify a single-case study.

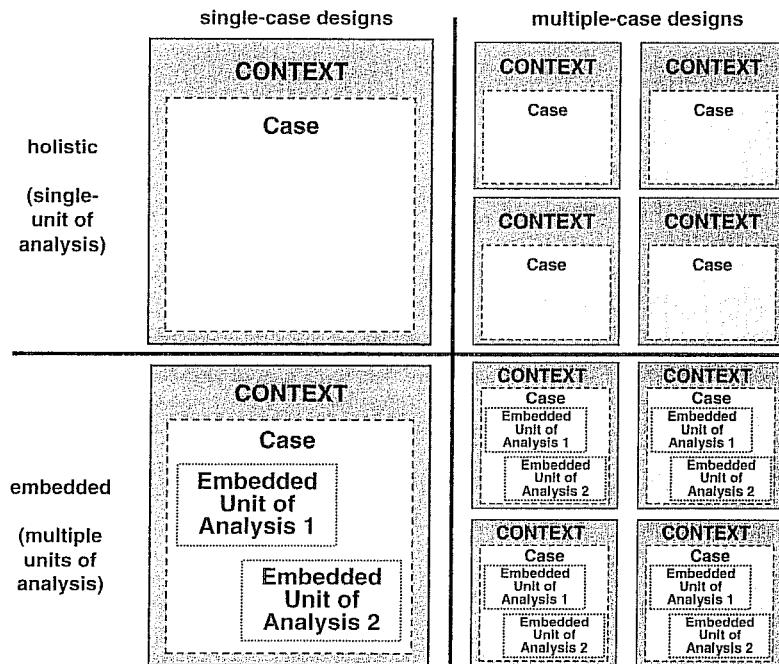


Figure 2.4 Basic Types of Designs for Case Studies
SOURCE: COSMOS Corporation.

One rationale for a single case is when it represents the *critical case* in testing a well-formulated theory (again, note the analogy to the *critical experiment*). The theory has specified a clear set of propositions as well as the circumstances within which the propositions are believed to be true. To confirm, challenge, or extend the theory, a single case may meet all of the conditions for testing the theory. The single case can then be used to determine whether a theory's propositions are correct or whether some alternative set of explanations might be more relevant. In this manner, like Graham Allison's (1971) comparison of three theories and the Cuban missile crisis (described in Chapter 1, BOX 1), the single case can represent a significant contribution to knowledge and theory building. Such a study can even help to refocus future investigations in an entire field. (See BOX 8 for another example, in the field of organizational innovation.)

A second rationale for a single case is when the case represents an *extreme* case or a *unique* case. Either of these situations commonly occurs

BOX 8

The Single Case Study as the Critical Case

One rationale for selecting a single-case rather than a multiple-case design is that the single case represents the *critical test of a significant theory*. Neal Gross et al. (1971) used such a design by focusing on a single school in their book, *Implementing Organizational Innovations*.

The school was selected because it had a prior history of innovation and could not be claimed to suffer from "barriers to innovation." In the prevailing theories, such barriers had been prominently cited as the major reason that innovations failed. Gross et al. showed that in this school, an innovation also failed but that the failure could not be attributed to any barriers. Implementation processes, rather than barriers, appeared to account for the outcomes.

In this manner, the book, though limited to a single case, represents a watershed in innovation theory. Prior to the study, analysts had focused on the identification of barriers; since the study, the literature has been much more dominated by studies of the implementation process.

in clinical psychology, in which a specific injury or disorder may be so rare that any single case is worth documenting and analyzing. For instance, one rare clinical syndrome is the inability of certain clinical patients to recognize familiar faces. Given visual cues alone, such patients are unable to recognize loved ones, friends, pictures of famous people, or (in some cases) their own image in a mirror. This syndrome appears to be due to some physical injury to the brain. Yet the syndrome occurs so rarely that scientists have been unable to establish any common patterns (Yin, 1970, 1978). In such circumstances, the single-case study is an appropriate research design whenever a new person with this syndrome—known as prosopagnosia—is encountered. The case study would document the person's abilities and disabilities, not only to determine the precise nature of the face recognition deficit but also to ascertain whether related disorders exist.

Conversely, a third rationale for a single case is the *representative* or *typical* case. Here, the objective is to capture the circumstances and conditions of an everyday or commonplace situation. The case study may represent a typical "project" among many different projects, a manufacturing firm believed to be typical of many other manufacturing firms in the same industry, a typical urban neighborhood, or a representative school, as examples. The lessons learned from these cases are assumed to be informative about the experiences of the average person or institution.

A fourth rationale for a single-case study is the *revelatory* case. This situation exists when an investigator has an opportunity to observe and analyze a phenomenon previously inaccessible to scientific investigation, such as Whyte's (1943/1955) *Street Corner Society*, previously described in Chapter 1, BOX 2. Another example is Elliot Liebow's (1967) famous case study of unemployed men, *Tally's Corner* (see BOX 9). Liebow had the opportunity to meet the men in one neighborhood in Washington, D.C., and to learn about their everyday lives. His observations of and insights into the problems of unemployment formed a significant case study because few social scientists had previously had the opportunity to investigate these problems, even though the problems were common across the country (as distinguished from the rare or unique case). When other investigators have similar types of opportunities and can uncover some prevalent phenomenon previously inaccessible to scientists, such conditions justify the use of a single-case study on the grounds of its revelatory nature.

A fifth rationale for a single-case study is the *longitudinal* case: studying the same single case at two or more different points in time. The theory of interest would likely specify how certain conditions change over time, and the desired time intervals to be selected would reflect the presumed stages at which the changes should reveal themselves.

These five rationales serve as major reasons for conducting a single-case study. There are other situations in which the single-case study may be used as a pilot case that is the first of a multiple-case study. However, in these latter instances, the single-case study cannot be regarded as a complete study on its own.

Whatever the rationale for doing single cases (and there may be more than the five mentioned here), a potential vulnerability of the single-case design is that a case may later turn out not to be the case it was thought to be at the outset. Single-case designs therefore require careful investigation of the potential case to minimize the chances of misrepresentation and to maximize the access needed to collect the case study evidence. A fair warning is not to commit yourself to the single case until all of these major concerns have been covered.

Holistic versus embedded case studies. The same case study may involve more than one unit of analysis. This occurs when, within a single case, attention is also given to a subunit or subunits (see BOX 10). For instance, even though a case study might be about a single organization, such as a hospital, the analysis might include outcomes about the clinical services and staff employed by the hospital (and possibly even some quantitative analyses based on the employee records of the staff). In an evaluation

BOX 9

The Revelatory Case as a Single Case

Another rationale for selecting a single-case rather than a multiple-case design is that the investigator has access to a situation previously inaccessible to scientific observation. The case study is therefore worth conducting because the descriptive information alone will be revelatory.

Such was the situation in Elliot Liebow's (1967) sociological classic, *Tally's Corner*. The book is about a single group of men living in a poor, inner-city neighborhood. By befriending these men, the author was able to learn about their lifestyles, their coping behavior, and in particular their sensitivity to unemployment and failure. The book provided insights into a subculture that has prevailed in many U.S. cities for a long period of time, but one that had been only obscurely understood. The single case showed how investigations of such topics could be done, thus stimulating much further research and eventually the development of policy actions.

BOX 10

An Embedded, Single-Case Design

Union Democracy (1956) is a highly regarded case study by three eminent academicians—Seymour Martin Lipset, Martin Trow, and James Coleman. The case study is about the inside politics of the International Typographical Union and involves several units of analysis (see the following table). The main unit was the organization as a whole, the smallest unit was the individual member, and several intermediary units also were important. At each level of analysis, different data collection techniques were used, ranging from historical to survey analysis.

study, the single case might be a public program that involves large numbers of funded projects—which would then be the embedded units. In either situation, these embedded units can be selected through sampling or cluster techniques (McClintock, 1985). No matter how the units are selected, the resulting design would be called an *embedded case study design* (see Figure 2.4, Type 2). In contrast, if the case study examined only the global nature of an organization or of a program, a *holistic design* would have been used (see Figure 2.4, Type 1).

Unit Being Characterized	Total System		Intermediate Units		Individuals	
	Issues, Data on Occupation; Union Laws; Policies; Historical Data; Convention Reports	Locals' Histories and Voting Records; Issues on Local Level; Size of Locals	Shops' Voting Records; Shop Size	Interviews with Leaders	Interviews of the Sample of Men	
ITU as a whole	Structural, environmental, behavioral properties	By inference, communication network (structural)				
Locals	Behavioral properties (militancy, etc.)	Behavioral properties, size	By inference, communication network (structural)	Structural, environmental, behavioral properties		
Shops			Behavioral properties, size		Distributions of individual properties	
Other immediate social environment of men	The social climate, by inference from dominant issues and election outcome	The social climate, by inference from dominant issues and election outcome			Chapel chairman's attributes; friends' attributes	
Men	By inference, dominant values and interests	By inference: values, interests, and loyalties (e.g., local over international)	By inference: values, interests, loyalties (e.g., to shop over local)	By inference: values	Behavior, background, values, attitudes	

SOURCE: Lipset, Trow, & Coleman (1956, p. 422). Reprinted by permission.

These two variants of single-case studies both have their strengths and weaknesses. The holistic design is advantageous when no logical subunits can be identified or when the relevant theory underlying the case study is itself of a holistic nature. Potential problems arise, however, when a global approach allows an investigator to avoid examining any specific phenomenon in operational detail. Thus, a typical problem with the holistic design is that the entire case study may be conducted at an abstract level, lacking any clear measures or data.

A further problem with the holistic design is that the entire nature of the case study may shift, unbeknownst to the researcher, during the course of study. The initial study questions may have reflected one orientation, but as the case study proceeds, a different orientation may emerge, and the evidence begins to address different research questions. Although some people have claimed such flexibility to be a strength of the case study approach, in fact, the largest criticism of case studies is based on this type of shift—in which the implemented research design is no longer appropriate for the research questions being asked (see COSMOS, 1983). Because of this problem, you need to avoid such unsuspected slippage; if the relevant research questions really do change, you should simply start over again, with a new research design. One way to increase the sensitivity to such slippage is to have a set of subunits. Thus, an embedded design can serve as an important device for focusing a case study inquiry.

An embedded design, however, also has its pitfalls. A major one occurs when the case study focuses only on the subunit level and fails to return to the larger unit of analysis. For instance, an evaluation of a program consisting of multiple projects may include project characteristics as a subunit of analysis. The project-level data may even be highly quantitative if there are many projects. However, the original evaluation becomes a project study (i.e., a multiple-case study of different projects) if no investigating is done at the level of the original case—that is, the program. Similarly, a study of organizational climate may involve individual employees as a subunit of study. However, if the data focus only on individual employees, the study will in fact become an employee and not an organizational study. In both examples, what has happened is that the original phenomenon of interest (a program or organizational climate) has become the context and not the target of study.

Summary. Single cases are a common design for doing case studies, and two variants have been described: those using holistic designs and those using embedded units of analysis. Overall, the single-case design is eminently justifiable under certain conditions—when the case represents (a) a critical test of existing theory, (b) a rare or unique circumstance, or (c) a

representative or typical case or when the case serves a (d) revelatory or (e) longitudinal purpose.

A major step in designing and conducting a single case is defining the unit of analysis (or the case itself). An operational definition is needed and some precaution must be taken—before a total commitment to the whole case study is made—to ensure that the case in fact is relevant to the issues and questions of interest.

Within the single case may still be incorporated subunits of analyses, so that a more complex—or embedded—design is developed. The subunits can often add significant opportunities for extensive analysis, enhancing the insights into the single case. However, if too much attention is given to these subunits, and if the larger, holistic aspects of the case begin to be ignored, the case study itself will have shifted its orientation and changed its nature. If the shift is justifiable, you need to address it explicitly and indicate its relationship to the original inquiry.

What Are the Potential Multiple-Case Designs (Types 3 and 4)?

The same study may contain more than a single case. When this occurs, the study has used a multiple-case design, and such designs have increased in frequency in recent years. A common example is a study of school innovations (such as the use of new curricula, rearranged school schedules, or new educational technology), in which individual schools adopt some innovation. Each school is the subject of an individual case study, but the study as a whole covers several schools and in this way uses a multiple-case design.

Multiple- versus single-case designs. In some fields, multiple-case studies have been considered a different “methodology” from single-case studies. For example, both anthropology and political science have developed one set of rationales for doing single case studies and a second set for doing what have been considered “comparative” (or multiple-case) studies (see Eckstein, 1975; George, 1979; Lijphart, 1975). This book, however, considers single- and multiple-case designs to be variants within the same methodological framework—and no broad distinction is made between the so-called classic (i.e., single) case study and multiple-case studies. The choice is considered one of research design, with both being included under the case study method.

Multiple-case designs have distinct advantages and disadvantages in comparison to single-case designs. The evidence from multiple cases is often considered more compelling, and the overall study is therefore regarded as being more robust (Herriott & Firestone, 1983). At the same

time, the rationale for single-case designs usually cannot be satisfied by multiple cases. The unusual or rare case, the critical case, and the revelatory case are all likely to involve only single cases, by definition. Moreover, the conduct of a multiple-case study can require extensive resources and time beyond the means of a single student or independent research investigator.

Therefore, the decision to undertake multiple-case studies cannot be taken lightly. Every case should serve a specific purpose within the overall scope of inquiry. Here, *a major insight is to consider multiple cases as one would consider multiple experiments*—that is, to follow a “replication” logic. This is far different from a mistaken analogy in the past, which incorrectly considered multiple cases to be similar to the multiple respondents in a survey (or to the multiple subjects within an experiment)—that is, followed a “sampling” logic. The methodological differences between these two views are revealed by the different rationales underlying the replication as opposed to sampling logics.

Replication, not sampling logic, for multiple-case studies. The replication logic is analogous to that used in multiple experiments (see Hersen & Barlow, 1976). For example, upon uncovering a significant finding from a single experiment, the immediate research goal would be to replicate this finding by conducting a second, third, and even more experiments. Some of the replications might have attempted to duplicate the exact conditions of the original experiment. Other replications might have altered one or two experimental conditions considered irrelevant to the original finding, to see whether the finding could still be duplicated. Only with such replications would the original finding be considered robust and worthy of continued investigation or interpretation.

The logic underlying the use of multiple-case studies is the same. Each case must be carefully selected so that it either (a) predicts similar results (a *literal replication*) or (b) predicts contrasting results but for predictable reasons (a *theoretical replication*). The ability to conduct 6 or 10 case studies, arranged effectively within a multiple-case design, is analogous to the ability to conduct 6 to 10 experiments on related topics; a few cases (2 or 3) would be literal replications, whereas a few other cases (4 to 6) might be designed to pursue two different patterns of theoretical replications. If all the cases turn out as predicted, these 6 to 10 cases, in the aggregate, would have provided compelling support for the initial set of propositions. If the cases are in some way contradictory, the initial propositions must be revised and retested with another set of cases. Again, this logic is similar to the way scientists deal with contradictory experimental findings.

An important step in all of these replication procedures is the development of a rich theoretical framework. The framework needs to state the conditions under which a particular phenomenon is likely to be found (a literal

replication) as well as the conditions when it is not likely to be found (a theoretical replication). The theoretical framework later becomes the vehicle for generalizing to new cases, again similar to the role played in cross-experiment designs. Furthermore, just as with experimental science, if some of the empirical cases do not work as predicted, modification must be made to the theory. Remember, too, that theories can be practical and not just academic.

For example, one might consider the initial proposition that an increase in using computers in school districts will occur when such a technology is used for both administrative and instructional applications, but not either alone. To pursue this proposition in a multiple-case study design, 3 or 4 cases might be selected in which both types of applications are present, to determine whether, in fact, computer use did increase over a period of time (the investigation would be predicting a literal replication in these 3 or 4 cases). Three or 4 additional cases might be selected in which only administrative applications are present, with the prediction being little increase in use (predicting a theoretical replication). Finally, 3 or 4 other cases would be selected in which only instructional applications are present, with the same prediction of little increase in use, but for different reasons than the administrative-only cases (another theoretical replication). If this entire pattern of results across these multiple cases is indeed found, the 9 to 12 cases, in the aggregate, would provide substantial support for the initial proposition. BOX 11 summarizes another example of a multiple-case replication design, but from the field of urban studies.⁵

This replication logic, whether applied to experiments or to case studies, must be distinguished from the sampling logic commonly used in surveys. The sampling logic requires an operational enumeration of the entire universe or pool of potential respondents and then a statistical procedure for selecting a specific subset of respondents to be surveyed. The resulting data from the sample that is actually surveyed are assumed to reflect the entire universe or pool, with inferential statistics used to establish the confidence intervals for which this representation is actually accurate. The entire procedure is commonly used when an investigator wishes to determine the prevalence or frequency of a particular phenomenon.

Any application of this sampling logic to case studies would be misplaced. First, case studies are not the best method for assessing the prevalence of phenomena. Second, a case study would have to cover both the phenomenon of interest and its context, yielding a large number of potentially relevant variables. In turn, this would require an impossibly large number of cases—too large to allow any statistical consideration of the relevant variables.

BOX 11

A Multiple-Case Replication Design

A common problem of the 1960s and 1970s was how to get good advice to city governments. Peter Szanton's (1981) book, *Not Well Advised*, reviewed the experiences of numerous attempts by university and research groups to collaborate with city officials.

The study is an excellent example of a multiple-case replication design. Szanton starts with eight case studies, showing how different university groups all failed to help cities. The eight cases are sufficient "replications" to convince the reader of a general phenomenon. Szanton then provides five more case studies, in which nonuniversity groups also failed, concluding that failure was therefore not necessarily inherent in the academic enterprise. Yet a third group of cases shows how university groups have successfully helped business, engineering firms, and sectors other than city government. A final set of three cases shows that those few groups able to help city government were concerned with implementation and not just with the production of new ideas, leading to the major conclusion that city governments may have peculiar needs in receiving advice.

Within each of the four groups of case studies, Szanton has illustrated the principle of *literal* replication. Across the four groups, he has illustrated *theoretical* replication. This potent case study design can and should be applied to many other topics.

Third, if a sampling logic had to be applied to all types of research, many important topics could not be empirically investigated, such as the following problem: Your investigation deals with the role of the presidency of the United States, and you are interested in studying the behavior of the incumbent from some leadership perspective. The leadership perspective, to be at all faithful to the complexity of reality, must incorporate dozens if not hundreds of relevant variables. Any sampling logic simply would be misplaced under such circumstances, as there have been only 43 presidencies since the beginning of the Republic. Moreover, you would probably not have the resources to conduct a full study of all the presidencies (and even if you did, you would still have too many variables in relation to the 43 data points available). This type of study just could not be done, following the sampling logic; if the replication logic is followed, however, the study is eminently feasible.

The replication approach to multiple-case studies is illustrated in Figure 2.5. The figure indicates that the initial step in designing the study must consist of theory development and then shows that case selection and

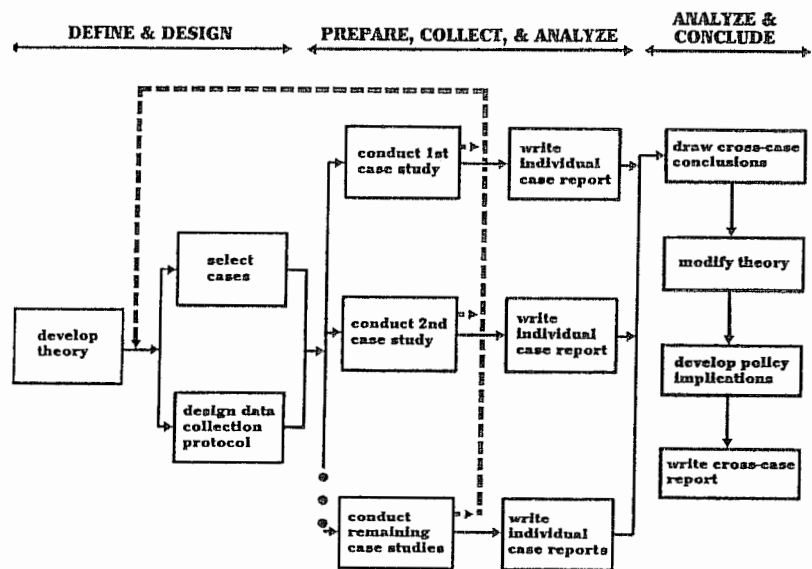


Figure 2.5 Case Study Method
SOURCE: COSMOS Corporation.

the definition of specific measures are important steps in the design and data collection process. Each individual case study consists of a “whole” study, in which convergent evidence is sought regarding the facts and conclusions for the case; each case’s conclusions are then considered to be the information needing replication by other individual cases. Both the individual cases and the multiple-case results can and should be the focus of a summary report. For each individual case, the report should indicate how and why a particular proposition was demonstrated (or not demonstrated). Across cases, the report should indicate the extent of the replication logic and why certain cases were predicted to have certain results, whereas other cases, if any, were predicted to have contrasting results.

An important part of Figure 2.5 is the dotted line feedback loop. The loop represents the situation in which important discovery occurs during the conduct of one of the individual case studies—for example, one of the cases did not in fact suit the original design. A second feedback loop (not shown) could represent the situation in which the discovery led to reconsidering one or more of the study’s original theoretical propositions. Under either circumstance, “redesign” should take place before proceeding further. Such redesign might involve the selection of alternative cases or changes

in the case study (i.e., data collection) protocol. Without such redesign, you risk being accused of distorting or ignoring the discovery, just to accommodate the original design. This condition leads quickly to a further accusation—that you have been selective in reporting your data to suit your preconceived ideas (i.e., the original theoretical propositions).

Overall, Figure 2.5 depicts a very different logic from that of a sampling design. The logic, as well as its contrast with a sampling design, may be difficult to follow and is worth extensive discussion with colleagues before proceeding with any case study design.

When using a multiple-case design, a further question you will encounter has to do with the *number* of cases deemed necessary or sufficient for your study. However, because a sampling logic should not be used, the typical criteria regarding sample size also are irrelevant. Instead, you should think of this decision as a reflection of the number of case replications—both literal and theoretical—that you need or would like to have in your study.

For the number of literal replications, an appropriate analogy from statistical studies is the selection of the criterion for establishing levels of significance: Much as the choice of “ $p < .05$ ” or “ $p < .01$ ” is not derived from any formula but is a matter of discretionary, judgmental choice, the selection of the number of replications depends on the certainty you want to have about your multiple-case results (as with the higher criterion for establishing statistical significance, the greater certainty lies with the larger number of cases). For example, you may want to settle for two or three literal replications when the rival theories are grossly different and the issue at hand does not demand an excessive degree of certainty. However, if your rivals have subtle differences or if you want a high degree of certainty, you may press for five, six, or more replications.

For the number of theoretical replications, the important consideration is related to your sense of the complexity of the realm of external validity. When you are uncertain whether external conditions will produce different case study results, you may want to articulate these relevant conditions more explicitly at the outset of your study and identify a larger number of cases to be included. For example, in studying neighborhood change, a common concern is that ethnically and racially different neighborhoods do not usually follow similar courses of change (e.g., Flippen, 2001). A multiple-case study of neighborhood change would therefore need to include at least some subgroups of cases that varied along ethnic or racial lines (and within each subgroup of cases, one would still want a minimum of two or three literal replications). In contrast, when external conditions are not thought to produce much variation in the phenomenon being studied, a smaller number of theoretical replications is needed.

Rationale for multiple-case designs. In short, the rationale for multiple-case designs derives directly from your understanding of literal and theoretical replications. The simplest multiple-case design would be the selection of two or more cases that are believed to be literal replications, such as a set of cases with exemplary outcomes in relation to some evaluation theory. Selecting such cases requires prior knowledge of the outcomes, with the multiple-case inquiry focusing on how and why the exemplary outcomes might have occurred and hoping for literal (or direct) replications of these conditions from case to case.

More complicated multiple-case designs would likely result from the number and types of theoretical replications you might want to cover. For example, investigators have used a “two-tail” design in which cases from both extremes (of some important theoretical condition, such as good and bad outcomes) have been deliberately chosen. Multiple-case rationales also can derive from the prior hypothesizing of different types of conditions and the desire to have subgroups of cases covering each type. These and other similar designs are more complicated because the study should still have at least two individual cases within each of the subgroups, so that the theoretical replications across subgroups are complemented by literal replications within each subgroup.

Multiple-case studies: Holistic or embedded. The fact that a design calls for multiple-case studies does not eliminate the variation identified earlier with single cases: Each individual case may still be holistic or embedded. In other words, a multiple-case study may consist of multiple holistic cases (see Figure 2.4, Type 3) or of multiple embedded cases (see Figure 2.4, Type 4).

The difference between these two variants depends on the type of phenomenon being studied and your research questions. In an embedded design, a study even may call for the conduct of a survey at each case study site. For instance, suppose a study is concerned with the delivery of services by different community mental health centers (Larsen, 1982). Each center may rightfully be the topic of a case study; the theoretical framework may dictate that nine such centers be included as case studies, three to replicate a direct result (literal replication) and six others to deal with contrasting conditions (theoretical replications).

For all nine centers, an embedded design is used because surveys of the centers’ clients (or, alternatively, examination of clients’ archival records) are needed to address research questions about the clients at the centers. However, the results of each survey will *not* be pooled across centers. Rather, the survey data will be part of the findings for each individual center, or case. These data may be highly quantitative, focusing on the attitudes and behavior of individual clients, and the data will be used along with

archival information to interpret the success and operations at the given center. If, in contrast, the survey data are pooled across centers, a multiple-case study design is no longer being used, and the investigation is likely to be using a survey rather than case study design.

Summary. This section has dealt with situations in which the same investigation may call for multiple-case studies. These types of designs are becoming more prevalent, but they are more expensive and time-consuming to conduct.

Any use of multiple-case designs should follow a replication, not a sampling logic, and an investigator must choose each case carefully. The cases should serve in a manner similar to multiple experiments, with similar results (a literal replication) or contrasting results (a theoretical replication) predicted explicitly at the outset of the investigation.

The individual cases within a multiple-case study design may be either holistic or embedded. When an embedded design is used, each individual case study may in fact include the collection and analysis of highly quantitative data, including the use of surveys within each case.

MODEST ADVICE IN SELECTING CASE STUDY DESIGNS

Now that you know how to define case study designs and are prepared to carry out design work, two pieces of advice may be offered.

Single- or Multiple-Case Designs?

The first word of advice is that although all designs can lead to successful case studies, when you have the choice (and resources), multiple-case designs may be preferred over single-case designs. Even if you can only do a “two-case” case study, your chances of doing a good case study will be better than using a single-case design. Single-case designs are vulnerable if only because you will have put “all your eggs in one basket.” More important, the analytic benefits from having two (or more) cases may be substantial.

To begin with, even with two cases, you have the possibility of direct replication. Analytic conclusions independently arising from two cases, as with two experiments, will be more powerful than those coming from a single case (or single experiment) alone. Second, the contexts of the two cases are likely to differ to some extent. If under these varied circumstances you still can arrive at common conclusions from both cases, they will have immeasurably expanded the external generalizability of your findings, again compared to those from a single case alone.

BOX 12

Two “Two-Case” Case Studies

12a. Contrasting Cases for Community Building

Chaskin (2001) used two case studies to illustrate contrasting strategies for capacity building at the neighborhood level. The author’s overall conceptual framework, which was the main topic of inquiry, claimed that there could be two approaches to building community capacity—using a collaborative organization to (a) reinforce existing networks of community organizations or (b) initiate a new organization in the neighborhood. After thoroughly airing the framework on theoretical grounds, the author presents the two case studies, showing the viability of each approach.

12b. Contrasting Strategies for Educational Accountability

In a directly complementary manner, Elmore, Abelmann, and Fuhrman (1997) chose two case studies to illustrate contrasting strategies for designing and implementing educational accountability—that is, holding schools accountable for the academic performance of their students. One case represented a lower-cost, basic version of an accountability system. The other represented a higher-cost, more complex version.

Alternatively, you may have deliberately selected your two cases because they offered contrasting situations, and you were not seeking a direct replication. In this design, if the subsequent findings support the hypothesized contrast, the results represent a strong start toward theoretical replication—again vastly strengthening the external validity of your findings compared to those from a single case alone (e.g., see BOX 12).

In general, criticisms about single-case studies usually reflect fears about the uniqueness or artifactual condition surrounding the case (e.g., special access to a key informant). As a result, the criticisms may turn into skepticism about your ability to do empirical work beyond having done a single-case study. Having two cases can begin to blunt such criticism and skepticism. Having more than two cases will produce an even stronger effect. In the face of these benefits, having at least two cases should be your goal. If you do use a single-case design, you should be prepared to make an extremely strong argument in justifying your choice for the case.

Closed Designs or Flexible Designs?

Another word of advice is that despite this chapter’s details about design choices, you should not think that a case study’s design cannot be modified by new information or discovery during data collection. Such revelations can be enormously important, leading to your altering or modifying your original design.

As examples, in a single-case study, what was thought to be a critical or unique case might turn out not to be so after initial data collection has started; ditto a multiple-case study, in which what was thought to be parallel cases for literal replication turn out not to be so. Under these discoveries, you have every right to conclude that your initial design needs to be modified. However, you should undertake any alterations only given a serious caution. The caution is to understand precisely the nature of the alteration: Are you merely selecting different cases, or are you also changing your original theoretical concerns and objectives? The point is that the needed flexibility should not lessen the rigor with which case study procedures are followed.

EXERCISES

1. *Defining the boundaries of a case study.* Select a topic for a case study you would like to do. Identify some basic questions to be answered by your case study. Does the naming of these questions clarify the boundaries of your case, with regard to the relevant length of time for which evidence is to be collected? The relevant organization or geographic area? The type of evidence that should be collected? The priorities for doing analysis?

2. *Defining the unit of analysis for a case study.* Examine Figure 2.6. Discuss each topic, possibly citing an example of a published case study on each topic. Understanding that each topic illustrates a different unit of analysis, do you think the more concrete units are easier to define than the more abstract ones? Why?

3. *Defining the criteria for judging the quality of research designs.* Define the four criteria for judging the quality of research designs: (a) construct validity, (b) internal validity, (c) external validity, and (d) reliability. Give an example of each type of criterion in a case study you might want to do.

4. *Defining a case study research design.* Select one of the case studies described in the BOXES of this book. Describe the research design of this case study. How did it justify the relevant evidence to be sought, given the basic research questions to be answered? What methods were used to draw conclusions, based on the evidence? Is the design a single- or multiple-case design? Is it holistic, or does it have embedded units of analysis?