Educational Research

An Introduction

Seventh Edition

✦

Meredith D. Gall
University of Oregon

Joyce P. Gall
University of Oregon

Walter R. Borg

Between the time Web site information is gathered and then published, it is not unusu-
al for some sites to have closed. Also, the transcription of URLs can result in unintend-
ed typographical errors. The publisher would appreciate notification where these errors
occur so that they may be corrected in subsequent editions.

Dedicated to All Our Students

Brief Contents

CONTENTS

2. To calculate the _____ of a measure, the measure is administered to a sample of individuals and then, after a delay, is again given to the same sample.
   a. alternate-form reliability
   b. test-retest reliability
   c. internal consistency
   d. reliability of subscores

3. Generalizability theory is useful for
   a. determining whether test-validity evidence generalizes to populations other than the ones on which it was collected.
   b. assessing the relative contribution of different sources of measurement error in a test.
   c. determining the degree of similarity across item characteristic curves.
   d. all of the above.

4. An important advantage of item response theory over classical test theory is that it
   a. speeds up the process of constructing a test.
   b. yields better validity generalizations.
   c. facilitates the construction of parallel tests of a construct.
   d. enables the measurement of more constructs in each item.

5. An important purpose of criterion-referenced measurement is to
   a. compare an individual's performance to the performance of other individuals.
   b. compare an individual's performance to an absolute standard of proficiency.
   c. extensively sample a broadly defined content domain.
   d. compare an individual's performance to his or her previous performance.

6. A good way to check the reliability of a performance assessment is to use
   a. alternate test forms
   b. generalizability theory.

c. hermeneutic theory.
d. item-response theory.

7. A potentially serious disadvantage of general personality inventories is that they usually are
   a. expensive to purchase.
   b. difficult to score.
   c. difficult to administer.
   d. based on self-report.

8. The *Test Locator* is designed specifically to
   a. identify tests that measure a particular construct.
   b. determine the publisher of a particular test.
   c. identify reviews of a particular test.
   d. serve as a refereed journal that identifies new tests and reviews them.

9. Computer-adaptive testing (CAT) differs most dramatically from traditional standardized tests with regard to
   a. the opportunity to randomize the ordering of item presentation.
   b. the feasibility of machine scoring.
   c. computer interpretation of the numerical test scores.
   d. the potential for each test-taker to take a different version of the test.

10. The likelihood of eliciting honest responses to personality measures can be increased by
    a. assuring individuals that their answers will be kept confidential.
    b. telling individuals that they can see their results on the personality measure after it has been scored.
    c. using computer-adaptive testing.
    d. using principles drawn from hermeneutics.

✦ Collecting Research Data with Questionnaires and Interviews

OVERVIEW

This chapter describes the use of questionnaires and interviews as data-collection instruments in both quantitative and qualitative research. We describe the distinctive characteristics of each method and compare their advantages and drawbacks. Also, we present techniques for constructing and administering these instruments.

OBJECTIVES

After studying this chapter, you should be able to

1. Describe the relative advantages and limitations of questionnaires and interviews in educational research.
2. Describe each step in constructing and administering a research questionnaire.
3. Describe several procedures for providing anonymity to questionnaire respondents.
4. Describe the relative advantages and limitations of closed-form and open-form items in questionnaires.
5. Explain the effects of respondents' knowledge and the number of items on attitude measurement in questionnaires.
6. Explain the advantages of precontacting a sample to whom a questionnaire will be sent.
7. Describe several features of a cover letter that are likely to increase the response rate to a mailed questionnaire.
8. Describe several strategies for following up with nonrespondents in order to maximize the response rate to a questionnaire.
9. Describe each step in preparing and conducting a research interview.

10. Describe the characteristics of key informant interviews, survey interviews, and focus group interviews.
11. Describe three levels of structure in quantitative research interviews and three levels of structure in qualitative research interviews.
12. Discuss the advantages of using telephone interviews and computer-assisted telephone interviewing.
13. Describe several factors that should be considered in selecting individuals to be research interviewers.
14. Describe the two phases in the training of research interviewers.
15. Describe four tasks that are involved in conducting an interview.
16. Explain the respective advantages and limitations of taking notes on an interview or recording it on tape or by computer.
17. Describe procedures for analyzing questionnaire or interview data using a quantitative research approach and using a qualitative research approach.

Questionnaires and Interviews as Data-Collection Methods

Questionnaires and interviews are used extensively in educational research to collect data about phenomena that are not directly observable: inner experience, opinions, values, interests, and the like. They also can be used to collect data about observable phenomena, but more conveniently than by direct observation. For example, it is much easier to use a questionnaire or interview to ask a principal how many teachers have at least one computer in their classroom than to walk around the school and make your own count. Of course, the advantage of ease is negated if the resulting data have poor validity.

Questionnaires are documents that ask the same questions of all individuals in the sample. (If the sample has subgroups, the questions asked of each subgroup may vary.) Respondents record a written or typed response to each questionnaire item. Also, the respondents typically control the data-collection process: They can fill out the questionnaire at their convenience, answer the items in any order, take more than one sitting to complete it, make marginal comments, or skip questions.

Interviews consist of oral questions asked by the interviewer and oral responses by the research participants. Interviews typically involve individual respondents, but there is increasing interest in conducting group interviews. Respondents typically speak in their own words, and their responses are recorded by the interviewer, either verbatim on audiotape or videotape, through handwritten or computer-generated notes, or in short-term memory for later note taking. The interviewer is largely in control of the response situation, scheduling with the participant a mutually agreeable time and place and then controlling the question pace and sequence to fit the circumstances of the situation.

Selecting between Questionnaires and Interviews

Questionnaires have two advantages over interviews for collecting research data: The cost of sampling respondents over a wide geographic area is lower, and the time required to collect the data typically is much less. Questionnaires, however, cannot probe deeply into respondents' beliefs, attitudes, and inner experience. Also, once the questionnaire has been distributed, it is not possible to modify the items, even if they are unclear to some respondents.

The major advantage of interviews is their adaptability. Skilled interviewers can follow up a respondent's answers to obtain more information and clarify vague statements. They also can build trust and rapport with respondents, thus making it possible to obtain information that the individual probably would not reveal by any other data-collection method.

Robert Jackson and J. W. M. Rothney did an extensive follow-up study of 890 adults five years after their high school graduation.[1] The entire sample was sent a four-page mailed questionnaire, and a subsample of 50 individuals was selected for a personal interview that included the same questionnaire items. The researchers found that 83 percent of the questionnaires were returned, whereas 98 percent of the planned interviews were completed. Two experienced counselors rated each questionnaire or interview protocol for evidence of personal problems. The mean number of problems yielded by the questionnaire data was 2.8, whereas the mean number of problems yielded by the interview data was 8.8. Thus the interview yielded more complete information, particularly information concerning negative aspects of the self.[2]

1. Jackson, R. M., & Rothney, J. W. M. (1961). A comparative study of the mailed questionnaire and the interview in follow-up studies. *Personnel and Guidance Journal, 39,* 569–571.
2. These findings are generally supported by another study: Legacy, J., & Bennett, F. (1979). A comparison of the mailed questionnaire and personal interview methods of data collection for curriculum development in vocational education. *Journal of Vocational Education Research, 4,* 27–39.

This advantage of the interview method is offset by some limitations. One is that it is difficult to standardize the interview situation so that the interviewer does not influence the respondent to answer questions a certain way. Another limitation is that interviews cannot provide anonymity for the respondents. In other words, the respondents must reveal their identity to the interviewer. Of course, the interviewer can analyze and report the interview data so that the identity of the participants is not revealed.

The questionnaire is more commonly used in quantitative research, because its standardized, highly structured design is compatible with this approach. The interview is more commonly used in qualitative research, because it permits open-ended exploration of topics and elicits responses that are couched in the unique words of the respondents. However, both methods can be used in either type of research. Robert Yin, for example, recommends using both methods when doing case study research.[3]

Validity and Reliability Issues

Questionnaires and interviews must meet the same standards of validity and reliability that apply to other data-collection measures in educational research. As we explain below, these standards are discussed at length elsewhere in the book. Therefore, do not interpret our brief treatment of validity and reliability here as an indication that they are tangential to good questionnaire and interview design.

If you are using a questionnaire or interview in a quantitative study, the validity and reliability standards described in the chapter on tests (Chapter 7) are relevant. For example, questionnaires often solicit respondents' opinions about particular topics and issues. If the researcher wishes to claim that these are the respondents' true opinions, she should collect evidence that the content of the items represents these constructs (content-related evidence of validity). Another option is to determine whether the respondents express similar opinions on other measures of the same construct (convergent evidence of validity).

In practice, researchers tend to apply looser validity and reliability standards to questionnaires and interviews than to tests because they typically are collecting information that is highly structured and likely to be valid (e.g., the respondents' years of schooling). Also, they are interested in the average response of the total group rather than the response of a single individual. A lower level of item reliability is acceptable when the data are to be analyzed and reported at the group level than at the level of individual respondents.

If you are using a questionnaire or interview in a qualitative study, the validity and reliability standards described in Chapter 14 are applicable. For example, the validity of a questionnaire or interview can be checked by using the method of triangulation described in that chapter.

Survey Research

The term *survey* frequently is used to describe research that involves administering questionnaires or interviews. The purpose of a survey is to use questionnaires or interviews to collect data from a sample that has been selected to represent a population to which the findings of the data analysis can be generalized. This emphasis on population generalization is characteristic of quantitative research, but not of qualitative research. Because we consider both qualitative and quantitative research in this chapter, we do not use the term *survey* as a general label for the use of questionnaires and interviews in research. However, the chapter includes references to publications about surveys where appropriate.

The term *survey research* occasionally is used as if it were a particular type of research design. We think that it is less confusing if this term is used to refer to research studies that

3. Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage.

rely primarily on questionnaires or interviews for data collection. The reason is that either type of instrument can be used to achieve the purposes of various research designs (particularly, descriptive, causal-comparative, and case-study designs). For example, Table 8.1 on page 236 presents the results of a data analysis that is typical of a causal-comparative research design. The results shown in Table 8.2 (see page 250) are typical of a descriptive research design.

Mangione, T. W. (1998). Mail surveys. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 399–427). Thousand Oaks, CA: Sage.

Steps in Constructing and Administering a Research Questionnaire

In this section we describe the major steps in carrying out a research study using a questionnaire: (1) defining research objectives, (2) selecting a sample, (3) designing the questionnaire format, (4) pretesting the questionnaire, (5) precontacting the sample, (6) writing a cover letter and distributing the questionnaire, (7) following up with nonrespondents, and (8) analyzing questionnaire data.

Step 1: Defining Research Objectives

Some researchers develop a questionnaire before they have thoroughly considered what they hope to obtain from the results. It is important that you define your research problem and list the specific objectives to be achieved, or hypotheses to be tested, by the questionnaire. You might start with a broad topic (e.g., teachers' involvement in staff development), but you should sharpen its focus before beginning on the design of the questionnaire.

D. A. deVaus suggested five types of questions that you can ask yourself for this purpose.[4] They are stated below in relation to the above-mentioned topic, teachers' involvement in staff development:

1. What is the *time frame* of your interest? Are you interested in teachers' current involvement in staff development, or do you want to study trends in their involvement over a period of years?
2. What is the *geographical location* of your interest? Do you want to study teachers in a particular state or region, or do you want to compare teachers in different locations?
3. Are you interested in a broad descriptive study or do you want to specify and compare different *subgroups*? For example, will you compare elementary, middle school, and high school teachers, or will you study teachers in general?
4. What *aspect* of the topic do you want to study? Are you interested in teachers' involvement in particular types of staff development activities, whether their involvement is mandatory or voluntary, or the amount of involvement over a given time period?
5. How *abstract* is your interest? For example, are you interested in reporting facts, or do you want to interpret the information, relate it to a broad social context, or develop theory from the findings?

In describing the steps involved in conducting a questionnaire study, we shall refer to a study by Corrine Glesne and Rodman Webb.[5] These researchers were interested in track-

---

4. deVaus, D. A. (1992). *Surveys in social research* (3rd ed.). Boston: Allen & Unwin.
5. Glesne, C., & Webb, R. (1993). Teaching qualitative research: Who does what? *International Journal of Qualitative Studies in Education, 6*, 253–266.

ing the growing emphasis on qualitative research in higher education in the United States. They wanted to determine who teaches qualitative research methods courses, the content of their courses, and their teaching methods. Their questionnaire was designed to obtain this information:

> The survey [questionnaire] asked about the training and academic background of qualitative research professors, content of courses, program requirements, and faculty perceptions of and interaction with students pursuing qualitative research dissertations.[6]

Glesne and Webb noted the irony of basing a study about the teaching of qualitative research methods courses on a quantitatively-oriented questionnaire survey. They chose to use questionnaires anyway because of their usefulness in collecting both closed and open-ended information from a widespread sample.

Step 2: Selecting a Sample

Once your research objectives or hypotheses are clearly stated, you should identify the target population from which your sample will be selected. (This and other sampling techniques are described in Chapter 6.) If you do not have thorough knowledge of the situation, you might make the mistake of sending your questionnaire to a group that does not have the desired information. For example, a graduate student seeking data on school financial policies sent questionnaires to principals of elementary and secondary schools. Many of the returned questionnaires were incomplete, and few specific facts of the sort wanted were obtained. This questionnaire failed because at that time the school superintendent and district specialists handled most matters concerning school finance. Because the principals who received the questionnaire had little specific knowledge about the topic, they were unable to supply the information requested.

The salience of the questionnaire content to the respondents (i.e., how important or prominent a concern it is for them) affects both the accuracy of the information received and the rate of response. A review of 181 studies using questionnaires judged to be "salient," "possibly salient," or "nonsalient" to the respondents revealed that the return rate averaged 77 percent for the salient studies, 66 percent for those judged possibly salient, and only 42 percent for those judged nonsalient.[7] These findings suggest the need to select a sample for whom your questionnaire items will be highly salient.

In the study by Glesne and Webb, the researchers gained access to a mailing list for the *International Journal of Qualitative Studies in Education,* which is a major journal publishing qualitative research studies. They then sent a copy of their questionnaire to 360 professors whose names were on the journal's mailing list. The researchers commented:

> This was, admittedly, a fishing-net approach. Our assumption was that this readership would include people who teach qualitative research methods courses, and not everyone on the list taught such courses.[8]

Using this admittedly biased sampling approach, they received usable questionnaires from 73 respondents in 37 different states. The sample included 40 men and 33 women. Twenty-five held the title of professor; 28 the title of associate professor, 18 the title of assistant professor, and 2 the title of lecturer.

---

6. Ibid., p. 254.
7. Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review, 43*, 447–462.
8. Glesne & Webb, Teaching qualitative research, p. 254.

Step 3: Designing the Questionnaire

Some research questionnaires appear to have been thrown together in an hour or two. The experience of receiving these haphazard questionnaires has led many educators to develop negative attitudes about the questionnaire as a research approach, and so they deposit them in the recycling box with little more than a quick glance. You will need to overcome these negative attitudes by careful construction and administration of your questionnaire. Figure 8.1 summarizes guidelines for designing questionnaires. These guidelines are based on research findings about factors that influence questionnaire return rate.

Guidelines for Designing a Questionnaire

1. Keep the questionnaire as short as possible.
2. Do not use technical terms, jargon, or complex terms that respondents may not understand.
3. Avoid using the words *questionnaire* or *checklist* on your form. Many persons are biased against these terms.
4. Make the questionnaire attractive by such techniques as using brightly colored ink or paper and laser printing.
5. Organize the items so they are easy to read and complete.
6. Number the questionnaire pages and items.
7. Put the name and address of the individual to whom the questionnaire should be returned both at the beginning and end of the questionnaire, even if a self-addressed envelope is included.
8. Include brief, clear instructions, printed in bold type and in upper and lower case (Words that are all capital letters are hard to read.)
9. Organize the questionnaire in a logical sequence. For example, you might group items with the same content or items having the same response options together.
10. When moving to a new topic, include a transitional sentence to help respondents switch their train of thought.
11. Begin with a few interesting and nonthreatening items.
12. Put threatening or difficult items near the end of the questionnaire.
13. Do not put important items at the end of a long questionnaire.
14. Provide a rationale for the items so that the respondent understands their relevance to the study.
15. Include examples of how to respond to items that might be confusing or difficult to understand.
16. Avoid terms like *several, most,* and *usually,* which have no precise meaning.
17. State each item in as brief a form as possible.
18. Avoid negatively stated items because they are likely to be misread by respondents. The negative word tends to be overlooked, and respondents might give an answer that is opposite to their real opinion.
19. Avoid "double-barreled" items that require the subject to respond to two separate ideas with a single answer. For example: *Although labor unions are desirable in most fields, they have no place in the teaching profession.*
20. When a general question and a related specific question are to be asked together, it is preferable to ask the general question first. If the specific question is asked first, it tends to narrow unnecessarily the respondent's focus when answering the general question that follows.
21. Avoid biased or leading questions. If the respondent is given hints as to the type of answer that is preferred, there is a tendency to give that response.

Source: Adapted from information in: Berdie, D. R., Anderson, J. F., & Niebuhr, M. A. (1986). Questionnaires: Design and use (2nd ed.). Metuchen, NJ: Scarecrow Press.

Anonymity

In most educational studies, respondents are asked to identify themselves, but anonymity might be necessary if highly personal or threatening information is requested. A questionnaire dealing with sexual behavior, for example, might receive more honest responses if the respondents remain anonymous.

The major problem with anonymous questionnaires is that follow-ups to improve the return rate are impossible. There are several solutions to this problem. One is to create a master code sheet that contains a code for each individual in the sample. The codes are put on the questionnaires. When an individual returns the questionnaire, the researcher can check off that person's name on the master code sheet. After a designated period of time, the researcher can determine which individuals have not returned their questionnaires and send them a new questionnaire.

This method is not completely anonymous, because the researcher can link the questionnaire (which has the code on it) to the individual's name by referring to the master code sheet. For complete anonymity, a variation of this approach can be used. The researcher sends each individual a prepaid postcard with the code on it and a questionnaire that contains no code. When the individual completes the questionnaire, she returns the questionnaire and the postcard separately. The postcard tells the researcher that this individual has completed the questionnaire, but he does not know which of the returned questionnaires belong to that individual.

Item Form

Writing items for questionnaires (and for interviews, too) may seem straightforward, but it is actually an art form. You need to be able to write succinctly and clearly. This is no easy matter. More importantly, you need to have a good understanding of your respondents so that you can use language that they understand, so that you can obtain all the information you need without exhausting their patience, and so that the items engage their interest and willingness to respond honestly.

A major difficulty in constructing questionnaire items is that educational terms often have multiple meanings. For example, the terms *charter school, standards-based education,* and *teacher empowerment* may mean different things depending on the individual educator and the region in which she works. If you use such a term in a questionnaire item, it is highly advisable to include a definition that corresponds to your research objectives. For example, suppose a researcher is interested in educators' responses to the charter-school movement, not as it is occurring nationally but within the state being studied. Given this objective, the item might read: "The state department of education adopted a statute in 2001 that allows school districts to start charter schools, which are defined as schools that receive district funding but are administered independently, albeit with mandatory conformance to standards of the state department of education. What is the current status of charter schools of this type in your district?"

A questionnaire item can be either closed form, meaning that the question permits only prespecified responses (similar to a multiple-choice question), or open form, meaning that respondents can make any response they wish (similar to an essay question). Which form to use is determined by the objective of the particular question. Evidence on the relative merits of closed and open questions, however, suggests that the two formats produce similar information.[9]

9. Bradburn, N. M. (1982). Question-wording effects in surveys. In: R. M. Hogarth (Ed.), *Question framing and response consistency* (pp. 65–76). San Francisco: Jossey-Bass.

The advantage of designing questions in closed form is that it makes quantification and analysis of results easier. For example, suppose you wish to know the size of a teacher's home town. Probably the least useful way to ask the question is: What is your home town? This question requires that you be able to read each teacher's response and then look it up in an atlas to determine the population. A somewhat better question would be: What is the population of your home town? In this case you could classify the responses into population categories such as those used by the U.S. Census Bureau. A still better approach would be to ask: What is the population of your home town? (Check one.), and provide the following response choices:

_____ rural, unincorporated
_____ incorporated, under 1000
_____ 1,000 to 2,500
_____ 2,500 to 5,000
_____ 5,000 to 10,000
_____ 10,000 to 50,000
_____ 50,000 to 250,000
_____ over 250,000
_____ don't know

This item requires little effort on your part to analyze the data, and also minimal effort from the respondents.

To determine the multiple-choice categories to use in closed-form questions, you can pilot-test the question by asking it in open form of a small number of respondents. Their answers can be used to develop the categories for the closed-form item. If you expect unusual responses, an "other" option can be provided.

In the questionnaire study on the teaching of qualitative research, Glesne and Webb began by interviewing several qualitative researchers about their training, teaching, and research. They used the interview information to develop an open-ended pilot questionnaire, and sent it to six professors of qualitative research. Feedback indicated that the open-ended questions were interesting, but time-consuming. There was a concern that few professors would take the hour or more needed to complete the questionnaire. Based on this feedback, the researchers redesigned the questionnaire into a closed-form format, with open-ended options attached to most items.

Measuring Attitudes

Questionnaires typically contain items each of which elicits a different bit of information. In effect, each item is a one-item test. The use of a one-item test is quite satisfactory when you are seeking a specific fact, such as number of years of full-time teaching experience, the number of wins and losses during a particular football coach's tenure, or the proportion of students failing intermediate algebra. When questions assess attitudes, however, the one-item test approach is questionable with respect to both validity and reliability. A questionnaire that measures attitudes generally must be constructed as an attitude scale and must use a substantial number of items (usually at least 10) in order to obtain a reliable assessment of an individual's attitude.[10]

If you are planning to collect information about attitudes, you should first do a search of the research literature to determine whether a scale suitable for your purposes already

10. For an expanded discussion of the measurement of attitudes by questionnaire, see: Schuman, H., & Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, wording, and context.* Thousand Oaks, CA: Sage.

has been constructed. (See Chapter 4 and Appendix E for information on locating such measures.) If a suitable scale is not available, you will need to develop one. Likert scales, which typically ask for the extent of agreement with an attitude item (for example, a five-point scale ranging from "strongly disagree" to "strongly agree") are a common type of attitude scale.

If you develop an attitude scale for your questionnaire study, you should pilot-test it in order to check its reliability and validity. Also, the pilot test should determine whether individuals in the sample have sufficient knowledge and understanding to express a meaningful opinion about the topic. Otherwise, their responses to the attitude scale will be of questionable value.

One method of dealing with respondents who lack familiarity with a topic is to include a "no opinion" option as one of the response alternatives for each attitude item. Even still, individuals with little or no information about the topic might express an opinion in order to conceal their ignorance, or because they feel social pressure to express a particular opinion. For example, Irving Allen conducted a questionnaire study of respondents attitudes' toward individuals and organizations that were the subject of considerable media attention at the time.[11] The respondents could express a favorable or unfavorable attitude using six Likert-type categories, or they could use a seventh category to express no knowledge of a particular individual or organization. Ten percent of the sample expressed a favorable or unfavorable attitude toward a fictitious organization, about which it was impossible for them to have any knowledge! The subjects responding to the fictitious item were found to have less formal education than the rest of the sample. They also were more likely to express attitudes toward the other organizations and individuals listed on the questionnaire than to check the "don't know" category, and to express more favorable attitudes.

As we stated above, a "no opinion" option for each attitude item might alleviate the problem identified in Allen's study. Another strategy is to include several information questions at the beginning of an attitude questionnaire that can be used to screen out respondents who display little or no knowledge of the topics being studied.

Web Questionnaires

Researchers increasingly are using the World Wide Web to administer questionnaires. For example, Mike Carbonaro and Joyce Bainbridge administered a Web questionnaire to obtain data about elementary teachers' use of Canadian children's literature in the classroom.[12] Of the 63 school districts in the target population, 53 agreed to participate. In the next phase of sample selection, the principal of each of 945 schools in the 53 districts was invited to participate; a total of 207 accepted. Each principal asked an appropriate teacher to complete the Internet-based questionnaire. A total of 110 teachers (out of a possible 207 teachers) completed it.

The following are distinctive features of the questionnaire design and administration process used by the researchers:

1. Teachers logged onto the survey Web site using a designated ID and password to avoid having any inappropriate person complete the questionnaire. Also, by logging on, the teachers acknowledged their consent to participate in the study.

✦ Touchstone in Research

Schmidt, W. C. (1997). World-Wide Web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, and Computers, 29,* 274–279.

11. Allen, I. L. (1966). Detecting respondents who fake and confuse information about question areas on surveys. *Journal of Applied Psychology, 50,* 523–528.
12. Carbonaro, M., & Bainbridge, J. (2000). Design and development of a process for web-based survey research. *Alberta Journal of Educational Research, 46,* 392–394.

2. Teachers responded to Likert scale items and closed-form items by clicking on "radio buttons" (a Web-page feature). They responded to rank-order items by entering a number and to open-form items by typing a response.
3. After completing the questionnaire, teachers clicked a SUBMIT button, which transmitted their data to the researchers' Web server. If a teacher clicked this item without completing the entire questionnaire, the Web software informed the teacher of which items still required completion.
4. The questionnaire data were secured in the researchers' Web server, so they were only available to them and the Web-server programmer.
5. Because the raw data were in electronic form, it was possible to import them directly into statistical software for analysis.

This approach to questionnaire design and administration has obvious advantages over conventional paper-and-pencil mailed questionnaires: Postal costs are eliminated; the possibility of missing data within questionnaires is eliminated; and there is no need to transfer data manually from the questionnaire into an electronic format and check for possible errors in the transfer process. Also, Internet questionnaires can be designed to be interactive: Items can be tailored to the individual respondent, and respondents can be given feedback as they complete the items.

Web questionnaires are a powerful research tool, but they have costs and limitations that you should consider in deciding whether to use one in your research study. You will need to have access to a Web server and the ability to use specialized software to design the questionnaire, to process incoming data, and to guard against data-security breaches and multiple submissions from the same respondent or a submission from an individual not in the sample. Also, each respondent needs to have access to a Web browser and the ability to use it. Otherwise, the research study is vulnerable to sampling bias.[13]

Step 4: Pilot-Testing the Questionnaire

You should carry out a thorough pilot test of the questionnaire before using it in your study. The pilot test should include a sample of individuals from the population from which you plan to draw your respondents. Also, the pilot-test form of the questionnaire should provide space for respondents to make criticisms and recommendations for improving the questionnaire. Another useful pilot-test strategy is to ask respondents to state in their own words what they think each question means. The questions should be revised and retested until they are understood accurately by all or most members of the pilot-test sample.

In a study of this procedure, William Belson elicited answers from respondents to 29 questions that incorporated problems of interpretation frequently found in questionnaire items.[14] He then studied the respondents' interpretations of the questions in a second in-depth interview. On average, only 29 percent of the respondents interpreted the questions within permissible limits of the intended interpretation. This finding demonstrates the importance of questionnaire wording and the need to check it by a pilot test.

Questionnaires mailed to educators generally can be expected to yield a higher percentage of replies than questionnaires mailed to samples of the general population. The response rate is higher for an educational questionnaire because it usually is targeted at a homogeneous group, and this makes it possible to prepare a specific appeal for participation that is likely to be effective. If you have received responses from less than 66 percent

13. Bradley, N. (1999). Sampling for Internet surveys. An examination of respondent selection for Internet research. *Journal of the Market Research Society, 41,* 387–395.
14. Belson, W. A. (1981). *The design and understanding of survey questions.* Lanham, MD: Lexington.

of the pilot-test sample, you probably should make changes in the questionnaire or in the procedures for administering it before sending the questionnaire to the participants in your main study.

Apparently a pilot test of the questionnaire used by Glesne and Webb was not conducted. The authors state:

> Our interview and subsequent survey questions grew out of our own experiences and from reading the few available sources on teaching qualitative research. We realize now that we should have gathered more demographic data such as information on ethnicity, salary, years of service, and years at current rank. . . . If we had asked for more information about respondents, we would likely have had other questions as well.[15]

The researchers' comments reinforce the desirability of pilot-testing a questionnaire before distributing it.

Step 5: Precontacting the Sample

Researchers have found that contacting respondents before sending a questionnaire increases the rate of response. A precontact involves the researchers identifying themselves, discussing the purpose of the study, and requesting cooperation. The precontact can take the form of a letter, postcard, or telephone call, but some evidence suggests that telephone contacts are the most effective.[16] Respondents also can be asked to return a postcard mailed to them indicating their willingness to cooperate.

Precontacts probably are effective because they alert respondents to the imminent arrival of the questionnaire, thus reducing the chance that it will be thrown out as junk mail. Precontacts also put a more personal, human face on the research study. Finally, having once agreed to cooperate, the respondent is under some psychological pressure to do so when the questionnaire arrives.

Step 6: Writing a Cover Letter

The main objective in doing a questionnaire survey is to get a high return rate. We know of studies where the return rate was as low as 20 percent, which makes it virtually impossible to generalize from the sample's data to the population that it is intended to represent.

Because the cover letter accompanying the questionnaire strongly influences the return rate, it should be designed carefully. The letter should be brief, but it must convey certain information and impressions. The purpose of the study should be explained so as to persuade the respondents that the study is significant and that their answers are important. When using a questionnaire that includes sensitive or potentially threatening questions, you should provide a specific description of how confidentiality will be maintained. You also should explain the conditions that you have established for informed consent (see Chapter 3). A sample cover letter is shown in Figure 8.2.

Subtle flattery in the cover letter can have a positive effect. If appropriate, you can emphasize the importance of the respondent's professional affiliation and the value of information that only members with this affiliation can supply. An offer to send the respondent a copy of the results also is effective. If such a promise is made, it should be honored. Failure to do so is unethical, and will lessen respondents' willingness to participate in other research studies.

15. Glesne & Webb, Teaching qualitative research, pp. 254–255.
16. Linsky, A. S. (1975). Stimulating responses to mailed questionnaires: A review. *Public Opinion Quarterly, 39,* 82–101.

Sample Cover Letter for a Mail Questionnaire

Letterhead paper	OKLAHOMA STATE UNIVERSITY Collegetown, Oklahoma M. A. Brown, President
	College of Education I. B. Smith, Dean
	February 1, 2002
Use mail-merge feature of word processor to personalize address →	Mr. A. B. Jones Superintendent of Schools Mediumtown, Oklahoma
Duplicated using word processor or offset process to look like individually typed letter →	Dear Mr. Jones:

Purpose of study →   The attached survey instrument concerned with procedures used in selecting elementary school principals is part of a statewide study being carried on cooperatively by the State Department of Public Instruction and Oklahoma State University. This project is concerned specifically with determining the present status of principal selection in our state. The results of this study will help to provide criteria to be used for developing *Importance of study* → better selection procedures and for improving the administrator training program at Oklahoma University.

Importance of respondent →   We are particularly desirous of obtaining your responses because your experience in principal selection will contribute significantly toward solving some of the problems we face in this important area of education. The enclosed instrument has been tested with a sampling of school administrators, and we have revised it in order to make it possible for us to obtain all necessary data while requiring a minimum of your time. The *Reasonable but specific time limit* → average time required for administrators trying out the survey instrument was 9.5 minutes.

Make return convenient for respondents →   We will appreciate it if you will complete the enclosed form prior to February 10th and return it in the stamped, self-addressed envelope enclosed. Other phases of this research cannot be carried out until we complete analysis of the survey data. We would welcome any com-*Assurance of confidentiality* → ments that you may have concerning any aspect of principal selection not covered in the instrument. Your responses will be held in strictest confidence.

Informed consent →   Informed consent procedures for this study are described on the enclosed sheet. Please take a moment now to read it.

Offer results →   We will be pleased to send you a summary of the survey results if *Thank respondent* → you desire. Thank you for your cooperation.

Print in different color to appear personally signed →   Sincerely yours,

                                        I. B. Smith

Signed by high-status educator →   I. B. Smith, Dean

                    Enclosure
                    sjc

It is desirable to associate your study with a professional organization with which prospective respondents might identify. For example, superintendents within a particular state might respond favorably to a letter signed by the state school superintendent or the president of a school administrators' association. If your study is well designed and deals with a significant problem, it usually is possible to have your cover letter signed by an individual whose endorsement represents a favorable symbol of authority to the respondents.

Responses to a questionnaire that is not aimed at a specific professional group tend to be more difficult to obtain because specific appeals cannot be made. Even with a heterogeneous group, however, you might be able to phrase your appeal in terms of common values that you expect most individuals to have, such as the importance of education and community improvement.

Researchers have explored the effect of enclosing a small cash reward with a questionnaire.[17] Such rewards, usually ranging from a quarter to a dollar, consistently increased the response rate, as did small gifts or premiums. Because most of these studies were conducted over 20 years ago, inflation might have weakened the effect of small rewards. A more recent study that offered a reward of two dollars to complete a 25-page questionnaire got a quicker reply from persons offered the reward, but the eventual response rate after follow-up attempts was about the same for respondents who received and who did not receive the reward.[18] If you provide a reward, it should be described as a token of appreciation rather than as payment for the respondent's time.

One of the items needed in the cover letter is a request that the questionnaire be returned by a particular date. Set this date so that the respondent will have sufficient time to fill out and return the questionnaire without rushing, but will not put it aside to do later. People tend to procrastinate if too generous a time allowance is given. A rule of thumb is to calculate the probable mailing time and allow the individual an additional week to complete the questionnaire and return it. You should include a stamped, self-addressed envelope with the questionnaire, so that individuals can respond with a minimum of inconvenience.

The design and neatness of your questionnaire and accompanying letter can improve the response rate. More expensive methods of duplication are usually worth the extra cost. A cover letter reproduced by the offset process on letterhead paper and signed with a different color ink will command more attention than one poorly dittoed on cheap paper. A word processor can produce individually typed letters, differing only in the names and addresses of the recipients, at a relatively low cost. Such letters are superior to the best offset copies and have the added advantage that small changes can be made at a reasonable cost in each letter to make it more individualized.

Step 7: Following Up with Nonrespondents

A few days after the time limit specified in the cover letter, it is desirable to contact nonrespondents by sending a follow-up letter, along with another copy of the questionnaire and another self-addressed envelope.[19] Because your original cover letter did not succeed with the nonrespondent group, there is little point in sending the same letter again. Instead, you should try a different approach in your appeal for cooperation. For example, if you used a

---

17. For reviews of studies offering cash rewards and premiums, see: Linsky, Stimulating responses; Heberlein & Baumgartner, Factors affecting response rates.

18. Shackelton, V. J., & Wild, J. (1982). Effects of incentives and personal contact on response rate to a mailed questionnaire. *Psychological Reports, 50,* 365–366.

19. Heberlein, T. A., & Baumgartner, R. (1981). Is a questionnaire necessary in a second mailing? *Public Opinion Quarterly, 45,* 102–108.

personal appeal in the initial letter, you might try a professional appeal in the first follow-up letter.

Successful follow-up letters usually take the approach that the researcher is confident the individual wished to fill out the questionnaire, but perhaps because of some oversight or an error on the researcher's part, it was overlooked. The follow-up letter then should repeat the importance of the study and the value of the individual's contribution, but with somewhat different language and emphasis from that in the original letter.

Postcard reminders have been tried also, and in some cases they have been found as effective as letters. However, Blaine Worthen and E. J. Brezezinski found that a form letter with another copy of the questionnaire obtained up to 7 percent more responses than a postcard with the same message.[20]

Figure 8.3 shows the pattern of responses reported in a review of 98 experimental studies on this problem. Although the reviewers point out that the results varied considerably across studies, these average percentages suggest what can be expected from different numbers of follow-ups. A few of the studies used four or more follow-ups, but this did not lead to a significant increase in returns over three follow-ups.

Suppose that you have a substantial percentage of nonrespondents after reaching the cutoff date for the return of the questionnaires to be included in your data analysis. You

---

A Synthesis of Research Findings on Response Rates for Initial Mailing and Different Numbers of Follow-ups
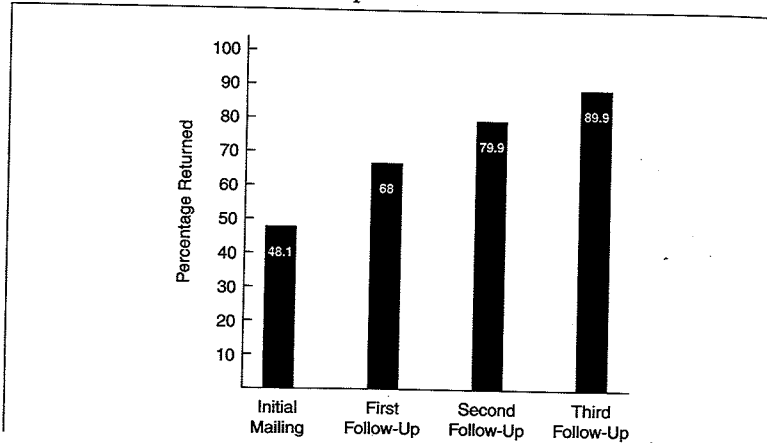
Source: Adapted from table 1 on p. 451 in: Heberlein, T.A., & Bumgartner, P. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review, 43,* 447–462.

---

20. Worthen, B. R., & Brezezinski, E. J. (1973, February). An experimental study of techniques for increasing return rate in mail surveys. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

---

should ask yourself: How would the results differ if all respondents had returned the questionnaire? If only a small percentage of respondents did not respond, this question is not critical. If more than 20 percent are missing, however, you need to pose this question because the sample for whom data are available might no longer be representative of the population to which you wish to generalize your findings. Researchers have found that respondents and nonrespondents to questionnaires do not differ in most personality characteristics, but nonrespondents tend to have achieved less academic success than respondents.[21]

The ideal method to determine whether nonrespondents to your questionnaire differ from the respondents is to randomly select a small number of individuals from the nonresponding group. Then solicit their cooperation in letting you administer the questionnaire to them in an in-person or telephone-interview format. Individuals who are reluctant to complete a questionnaire may be more amenable to this approach.

A sample of 20 individuals should be sufficient to check the nonresponding group. A comparison of their responses to each item with the responses of those who replied initially will enable you to determine whether the nonresponding sample is biased. In this case, you should note these differences and discuss their significance in reporting the results of the responding sample.

Step 8: Analyzing Questionnaire Data

The researchers who studied the teaching of qualitative research in institutions of higher education in the United States followed a typical approach to analyzing questionnaire data:

> All forced-choice answers in the survey were coded and entered into the *Ecstatic* analysis program for quantitative data. This procedure allowed the easy generation of percentages, means, ranges, and cross tabulations. All comments and open-ended answers were entered in full into the *Ethnograph* text analysis program which assisted in coding and sorting respondents' words so that patterns could be ascertained.[22]

The quantitative data were analyzed to yield frequencies and percentages of respondents checking each response category on particular closed-ended questions. For example, Table 8.1 shows the degree to which professors teaching qualitative research methods courses themselves have had formal coursework in qualitative research. Glesne and Webb concluded from these results that

> new faculty responsible for teaching qualitative methods courses generally have had more formal course work in qualitative methodology than older professors. . . . These percentages are not surprising, since the offering of qualitative research courses is fairly recent in most colleges of education. . . . What is surprising is that at each of the professorate levels, a higher percentage of women were trained in qualitative research than men.[23]

The results shown in Table 8.1 are consistent with causal-comparative research design (see Chapter 10). None of the three variables—professorial rank, gender, and amount of qualitative-research coursework—was experimentally manipulated by the researchers. The three variables were related to each other in the data analysis in such a way as to reveal

---

21. For a summary of these findings, see the discussion of differences between volunteers and nonvolunteers in Chapter 6.
22. Glesne & Webb, Teaching qualitative research, p. 254. *Ecstatic* is a computer program for statistical analysis of data. *Ethnograph* is a computer program for analyzing documents and transcriptions. Cross-tabulations are statistical analyses that show the relationship between two variables.
23. Glesne & Webb, Teaching qualitative research, p. 255.

Formal Coursework in Qualitative Research Taken by Faculty Teaching Qualitative Methods Courses

Title	Coursework (%)	Little or No Coursework (%)	Total
Full Professors	6 (24%)	19 (76%)	25
Males*	2 (8%)	17 (68%)	19
Females	4 (16%)	2 (8%)	6
Associate Professors	16 (57%)	12 (43%)	28
Males	5 (18%)	7 (25%)	12
Females	11 (39%)	5 (18%)	16
Assistant Professors	15 (83%)	3 (16%)	18
Males	6 (33%)	1 (5%)	7
Females*	9 (50%)	2 (11%)	11

*Indicates that one respondent did not answer the question; the two lecturers were not included in the table.
Source: Table 1 on p. 255 in: Glesne, C., & Webb, R. (1993). Teaching qualitative research: Who does what? *International Journal of Qualitative Studies in Education, 6,* 253–266. Reprinted with permission of Taylor & Francis, Ltd.

potential cause-and-effect relationships. In particular, the observed relationship between gender and coursework suggests (but by no means *proves*) that gender plays a causal role in whether an individual chooses to specialize in qualitative research methods. Further research would be necessary to determine whether and how these two variables are causally related to each other.

We emphasize this feature of Glesne and Webb's data analysis because it is commonly assumed that questionnaires and interviews are only suitable, or most suitable, for descriptive research. In fact, questionnaires and interviews can be used in various research designs.

Glesne and Webb included several comments from respondents in response to particular questions. In this way, the reader is provided an emic perspective, that is, the respondents' perspective on the phenomenon being studied. For example, they included this comment from a respondent to an open-form question about what excites them about teaching qualitative research methods courses:

> I really enjoy and feel challenged to support students' research and their forays into what is often new territory—seeing their eyes light up when they see what research can be and when they move from "I thought it'd be easier" to "Wow, I never knew what was involved!"[24]

Quantitative data collected by questionnaires can be analyzed by the statistical methods described in Chapter 5. Methods for analyzing qualitative data are described in Chapter 14.

Steps in Preparing and Conducting Research Interviews

The steps involved in using interviews in educational research are similar to those involved in using questionnaires. The steps are (1) defining the purpose of the study, (2) selecting a

24. Ibid., p. 262.

sample, (3) designing the interview format, (4) developing questions, (5) selecting and training interviewers, (6) doing a pilot test of the interview procedures, (7) conducting the interviews, and (8) analyzing the interview data.[25]

Step 1: Defining the Purpose of the Interview

The first step in a study that will employ interviews to collect research data is to define the purpose of the study. Your purpose will determine the nature of the interview because different purposes require different levels of structure, types of questions, and interviewer qualifications.

The different interests and orientations of researchers have given rise to different types of interviews. Several have been developed for a particular purpose and context, but you may be able to adapt them for investigating your research problem. The following are three major types of research interviews: key informant interviews, survey interviews, and group interviews.

Key Informant Interviews

In a key informant interview, the interviewer collects data from individuals who have special knowledge or perceptions that would not otherwise be available to the researcher. Key informants often have more knowledge, better communication skills, or different perspectives than other members of the defined population.

A study by Eleanor Lynch, Rena Lewis, and Diane Murphy illustrates the use of key informant interviews in educational research.[26] The researchers described the purpose of their study as follows:

> The changing needs of children with chronic illnesses pose some serious questions:
> - How can school systems respond most effectively to the needs of children with chronic illness?
> - Should children be served under special education? If yes, how can procedures be adapted to allow for the week-to-week differences in children's educational needs?
> - How can we ensure that adequate information is available to teachers and other school personnel working with children with chronic illnesses?
> - What do families want for their children with chronic illness and how can schools help support families' wishes?[27]

The researchers identified two groups of key informants from whom to collect interview data relating to these questions. One group was school district personnel who were in charge of services for children with chronic illnesses. This group could be expected to have expert knowledge about the research problem being investigated. The other group included parents of chronically ill children. The parents were key informants because they had direct knowledge of their family's needs with respect to the chronically ill child. We refer to this study by Lynch and her associates in the following sections to illustrate the various steps that are involved in using interviews in a research study.

Survey Interviews

The purpose of survey interviews is to supplement data that have been collected by other methods. Margaret LeCompte, Judith Preissle, and Renata Tesch describe three types of

25. These steps are adapted from Stewart, C. J., & Cash, W. B., Jr. (1997). *Interviewing: Principles and practices* (8th ed.). Madison, WI: Brown & Benchmark.
26. Lynch, E. W., Lewis, R. B., & Murphy, D. S. (1992). Educational services for children with chronic illnesses: Perspectives of educators and families. *Exceptional Children, 59,* 210–220.
27. Ibid., pp. 211–212.

survey interviews.[28] The first type is the **confirmation survey interview,** which is a structured interview that produces evidence to confirm earlier findings. These interviews are especially useful in large-scale questionnaire studies where in-depth interviewing cannot be carried out for all respondents.

The second type of survey interview is the **participant construct interview, which is used to learn how informants structure their physical and social world. The result is a set of category systems used by the participant. For example, LeCompte conducted a research study in which she asked kindergarten children to tell her all the things they thought they and their teachers could do in kindergarten.[29] The responses were used to develop a typology of children's perceptions of student and teacher roles.**

The third type of survey interview involves projective techniques. **Projective techniques** use ambiguous stimuli to elicit subconscious perceptions that cannot be observed in the natural setting or solicited through regular interviewing. Projective techniques are further explained in Chapter 7.

Focus Group Interviews

A group interview involves addressing questions to a group of individuals who have been assembled for this specific purpose. The individuals are selected because they are well informed about the research topic.

Group interviews have been used extensively by social science researchers and marketing researchers, who call them *focus group interviews,* or simply *focus groups.* Richard Krueger and Mary Anne Casey identified the following as characteristics of a **focus group:**

> [It is] a carefully planned discussion designed to obtain perceptions on a defined area of interest in a permissive, nonthreatening environment. It is conducted with approximately seven to ten people by a skilled interviewer. The discussion is relaxed, comfortable, and often enjoyable for participants as they share their ideas and perceptions. Group members influence each other by responding to ideas and comments in the discussion.[30]

Qualitative researchers have become interested in the use of focus groups to collect data in recent years. These researchers are finding that the interactions among the participants stimulate them to state feelings, perceptions, and beliefs that they would not express if interviewed individually. Also, the focus group technique avoids putting the interviewers in a directive role. They ask questions to initiate discussion, but then allow participants to take major responsibility for stating their views and drawing out the views of others in the group.

Procedures for selecting a focus group are described in the next section.

Step 2: Selecting a Sample

A sample of respondents should be selected using one of the quantitative or qualitative sampling techniques described in Chapter 6. To study needed services for children with chronic illnesses, Lynch and her associates selected two samples: (1) a stratified sample of school districts in California, and (2) a nonrandom sample of families with such children. Separate interview guides were developed for respondents in each sample.

28. LeCompte, M. D., Preissle, J., & Tesch, R. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). San Diego, CA: Academic Press.
29. LeCompte, M. D. (1980). The civilizing of children: How young children learn to become students. *The Journal of Thought, 15,* 105–126.
30. Krueger, R. A., & Casey, M. A. (2000). *Focus groups: A practical guide for applied research* (3rd ed.). Thousand Oaks, CA: Sage. Quote appears on p. 18.

Interviewers typically interview one respondent at a time, as in Lynch's study. It also is possible to conduct a focus group interview, as we explained in the preceding section. The focus group may consist of an established group, such as the teachers in a particular school. When using an established group, the researcher needs to be sensitive to pre-existing relationships among the group members. The focus group technique works best when all members are on an equal basis—for example, all the teaching staff of a preschool. If the school principal is included, the teachers may feel inhibited about sharing their actual perceptions of the phenomena being investigated.

Focus groups generally include seven to ten individuals. This group size encourages a wide sampling of views, but is not so large that some individuals do not have the opportunity to speak.

When interviewing individuals, you can arrange to meet with each respondent at your mutual convenience. In a focus group, however, all respondents must be assembled at the same time and place. This is not an easy task, and so you will need to follow systematic procedures to ensure that it is accomplished successfully.[31]

Step 3: Designing the Interview Format

In quantitative research, the interview generally is structured to expose all respondents to a nearly identical experience. Thus, the opening statement, interview questions, and closing remarks should be carefully specified in advance to ensure that data from all respondents can be compared meaningfully. In qualitative research, however, the interview format is not so tightly structured because the researcher's goal is to help respondents express their view of a phenomenon in their own terms.

Quantitative and qualitative research interviews also differ in whether the variables are prespecified. In quantitative studies, the variables of interest to the researcher generally are prespecified. For example, suppose the researcher wishes to determine the factors that influence students to choose a particular major in college. Through a review of the literature, the researcher might discover that parents, parents' friends, relatives, teachers, and other students are possible sources of influence. In designing the interview the researcher could ask questions about each of these sources of influence, for example: "Did your father influence your choice of a major?"

If a similar study were done from a qualitative research perspective, there might be little or no prespecification of variables. Instead, the interview questions might be broader in nature, for example: "How did you come to be an Economics major?" At the stage of analyzing the data, the researcher may choose to identify quantifiable variables or broad themes and patterns.

Interview Formats in Qualitative Research

Michael Patton describes three basic approaches to collecting qualitative data through open-ended interviews.[32] The three approaches, which are described below, vary in degree of structure.

The informal conversational interview relies entirely on the spontaneous generation of questions in a natural interaction, typically one that occurs as part of ongoing participant observation fieldwork. (Participant observation is explained in Chapter 9.) Because the conversation appears natural, the research participants may not even realize that they are being interviewed.

31. Procedures for arranging a focus group meeting are described in Chapter 6 of Krueger & Casey, *Focus groups.*
32. Patton, M. Q. (2001). *Qualitative evaluation and research methods* (3rd ed.). Thousand Oaks, CA: Sage.

The **general interview guide approach** involves outlining a set of topics to be explored with each respondent. The order in which the topics are explored and the wording of the questions are not predetermined. They can be decided by the interviewer as the situation evolves.

The **standardized open-ended interview** involves a predetermined sequence and wording of the same set of questions to be asked of each respondent in order to minimize the possibility of bias. This approach is particularly appropriate when several interviewers are used to collect data.

Interview Formats in Quantitative Research

Like qualitative research interviews, interviews in quantitative research vary in degree of structure. The three basic approaches are described below.

The **structured interview** involves a series of closed-form questions that either have yes-no answers or can be answered by selecting from among a set of short-answer choices. The respondents' answers are not followed up to obtain greater depth, and thus are similar to those obtained from a questionnaire. The advantage of an interview over a questionnaire in this case, however, is that the response rate can be increased because the interviewer can interact with individuals to reduce the number of unusable or "don't know" responses.

The **semistructured interview** involves asking a series of structured questions and then probing more deeply using **open-form questions** to obtain additional information. For example, suppose a researcher is investigating the relationship between students' high school experiences and their subsequent achievement in college. In one part of the interview, the interviewer might try to elicit significant experiences in coursework by asking all respondents: "What course did you like best?" Suppose the respondent answers, "I liked chemistry best because the teacher made it interesting." At this point, the interviewer might probe by asking: "How did the teacher make it interesting?" Another respondent might say, "I liked my government class because we talked about real-life problems." The interviewer then might probe by asking such questions as: "What are some examples of these problems?" and "Why did you find these problems interesting?" In these two examples, the interviewer began with the same initial question, but **asked different probing questions based on the respondent's answer.** This interview approach has the advantage of providing reasonably standard data across respondents, but of greater depth than can be obtained from a structured interview.

The study of children with chronic illness that we have been describing involved a semistructured telephone interview. When an appropriate respondent from each district was identified, this individual was sent an information packet that included the interview protocol. This procedure provided the respondent an opportunity to review the questions and prepare for the interview. Specific questions were drafted and formatted by the three researchers working as a team, reviewed and revised by State Department of Education personnel who had content and research expertise, and subjected to final review and revision by an advisory committee composed of experts or representatives from various constituencies (e.g., a special educator, a teacher, a parent of a chronically ill child, and a university student who had a chronic illness).

The **unstructured interview** does not involve a detailed interview guide. Instead, the interviewer asks questions that gradually lead the respondent to give the desired information. Usually the type of information sought is difficult for the respondent to express or is psychologically sensitive. For this reason the interviewer must adapt continuously to the respondent's state of mind. This format is highly subjective and time-consuming.

Telephone Interviews

The telephone commonly is used for interviewing because it is much less expensive than face-to-face interviews, especially when the sample is geographically dispersed. Although relatively low cost is the greatest advantage of telephone interviews, they have other significant advantages as well:

1. You can select respondents from a broader accessible population than if interviewers needed to travel to the location of each respondent.
2. Because all interviewers can work from a central location, monitoring of interviews and quality control is easier.
3. Little cost is incurred when no one answers, making frequent callbacks feasible.
4. Many groups, such as business people, school personnel, and parents, are easier to reach by telephone than by personal visits.
5. Telephone interviewing provides safe access to dangerous locations and access to restricted locations where interviewers might not be admitted.

There is some evidence that telephone interviews can be used to collect sensitive data. One study found that for nonthreatening questions respondents' distortions were slightly higher for telephone interviews than for face-to-face interviews.[33] For threatening questions, the reverse was true. Although it would seem easier to establish rapport in a face-to-face interview, the physical presence of the interviewer might increase the perceived threat of questions about sensitive topics. Hanging up a phone obviously is easier than ejecting an interviewer from one's home or office. Nevertheless, some investigators have been successful in completing a very high percentage of telephone interviews, even when dealing with sensitive topics. In one study, completed interviews were obtained from 74 percent of the sample in personal interviews and 70 percent in telephone interviews.[34] Because the same items were used for the personal and telephone interviews, it was possible to compare the responses for the two methods. The results generally were very similar over a wide range of topics and item formats.

When selecting a sample for a telephone interview, you will need the telephone number of each individual whom you select. If you use an organization's directory to select a sample, the members' telephone numbers may be listed. If not, you will need to determine the phone numbers by another procedure. The city telephone directory is useful, except for individuals who have unlisted numbers. (CD-ROMs and Web databases listing all telephone numbers in the United States are available as well.) Keep in mind that some individuals, especially those with low incomes, do not have telephones.

You should avoid eliminating individuals selected for your sample because their telephone number is unlisted or they have no telephone. To do so would create a biased sample, and thus weaken the generalizability of your research results.

As we noted above, the study by Lynch and her associates concerning the educational needs of children with chronic illnesses was conducted by telephone.

Computer-Assisted Telephone Interviews

Computer-assisted telephone interviews involve the use of a computer to assist in gathering information from telephone interviews. This method virtually eliminates two major

33. Graves, R. M., & Kahn, R. L. (1979) *Surveys by telephone: A national comparison with personal interviews.* New York: Academic Press.
34. Ibid.

sources of common errors in interviews, namely, recording data in the wrong place on the form and asking the wrong questions. Most telephone interviews require the interviewer to jump to a different part of the form depending on the response of the person interviewed. For example, if the question is "Are you employed?" a "yes" response might call for the interviewer to check this response on the interview guide and then turn three pages to a set of questions on mode of employment. If the interviewer does not turn the correct number of pages, inappropriate questions may be asked next, with the possible result of a badly shortened interview.

This problem can be avoided by developing a computer program that not only records the subject's responses, but also branches to the next question that should be asked. For example, as the interviewer types "yes" (or a code like "Y") into the computer, the response will be recorded and the computer can be programmed to jump three pages in the computer file containing the interview guide and display the first question on mode of employment. The interviewer does not have to worry about turning pages, nor does she even see any inappropriate questions. The next question that appears is the one needed. Response accuracy generally increases with such computer-assisted interview techniques, because the interviewer can concentrate on responses rather than worry about what question to ask next. Also, because the interviewees' responses are entered into a computer file while the interview is in progress, the data are ready for statistical analysis by computer as soon as all the interviews are completed.

Step 4: Developing Questions

Whether questions are developed in advance of the interview or during each interview itself depends mainly on the type of interview. The unstructured interview in quantitative research and the informal conversational interview in qualitative research involve on-the-spot formulation of questions, based on a general plan and the interviewer's reading of relevant characteristics of each respondent (e.g., level of poise, talkativeness, and intelligence). The other interview formats make greater use of prespecified closed-form and open-form questions. For example, in the study of the needs of children with chronic illnesses, a series of open-form and closed-form questions was used.

The formulation of good questions in interviews at the unstructured end of the continuum depends on the interviewer's ability to think on his feet during the interview process. Developing questions for more structured interviews is best done by designing and trying out an interview guide. An interview guide specifies the questions, the sequence in which they are to be asked, and guidelines for what the interviewer is to say at the beginning and end of each interview. The interview guide should list the response options for each closed-form question and provide space for the interviewer to write down answers that do not fit prespecified response categories.

Figure 8.4 shows an interview guide from a study by Michael Ann Rossi. Her research project involved case studies of "teacher-leaders" in mathematics, that is, teachers who had participated in a special mathematics institute and returned to their school district in a leadership role to improve mathematics instruction. Rossi was particularly interested in the strategies and skills of these teacher-leaders, and the outcomes that they effected. Her data-collection method involved interviewing each teacher-leader, the teacher-leader's supervisor, and teachers with whom the teacher-leader had worked. Separate interview guides were developed for each of these groups. The interview guide shown in Figure 8.4 was for interviewing teachers with whom the teacher-leader had worked.

A Guide for Interviewing Teachers Who Had Worked with a Teacher-Leader

(Start by alluding to introduction from teacher-leader.) This is a visit to get acquainted. *It's not an evaluation of you*, of your school program, or of the teacher-leader. I would like to get a picture of mathematics teaching and learning in your school. My main focus is how you have worked with other people along the way regarding getting help with mathematics teaching and learning. I want to understand the story since _____ until now. I have a number of specific questions to ask.

1. Background
   a. Name
   b. Job title (or role)
   c. Thumbnail sketch of your job; what you do, who you work with.
   d. How long have you been teaching?
   e. How long have you been in this school?
2. I'm interested in the *flavor* or feeling in the school.
   a. Can you give me 3 or 4 adjectives that would describe that?
   b. Can you think back to when your school first got involved with changing mathematics teaching and learning?
      • when was that?
      • why did the school get involved?
      • how did you personally get involved?
      • what did you expect?
      • what do you think the teacher-leader expected?
   c. Could you give me a quick sketch of how mathematics teaching and learning is changing in your school right now?
      • are teachers involved as individuals?
      • how many are involved?
      • what is the role of the principal? is *she* supportive of change, or blocking it?
      • what is the purpose of the change?
      • what procedures or guidelines are followed, methods used?
      • what does the teacher-leader do?
3. a. Describe your involvement since that time.
   b. What contact have you had with others who are involved? (Especially PROBE for communication, cooperation, peer coaching)
   c. Are there stages or phases that can be identified regarding *your* involvement with the change? Your *school's* involvement?
4. Generally speaking, what do you see as the teacher-leader's main role?
   a. What's been *her* main contribution to your school's mathematics program?
   b. Can you give me a few adjectives to describe *her style,* way of working with people?
   c. What do you see as *her* special strengths?
   d. Could you tell me about a specific incident when ___ was *especially helpful?*
      • What did *she* do?
      • Why did you think this was helpful?
      • What skills did you see *her* using in this situation?
   e. Now let's take another incident.
      • What did *she* do, in detail?
      • Why did you think this was helpful?
      • What skills did you see *her* using in this situation?
   f. Do you think *her* skills and strengths have changed since you've known her? (GET ILLUSTRATIONS AND EXAMPLES)
   g. Ask questions d, e, and f except use *practical* rather than helpful.

FIGURE 8.4   Continued

5. I'm interested in the program's results. For (1) you, (2) other teachers, and (3) the students:
   a. What results have occurred?
   b. Why do you think these results happened?
   c. In your opinion, how did ___ contribute to these results?
6. a. What would you say are the necessary ingredients of success in this kind of program?
   b. Specifically, what recommendations do you have for how Math Project teacher-leaders work and who is selected?
7. Do you have anything else to add?

Source: Appendix F in: Rossi, M. A. (1993). The California Mathematics Project: Empowering elementary teachers to be leaders and change agents in mathematics reform *Dissertation Abstracts International, 54* (09), 3314A. (UMI No. 9405218)

Step 5: Selecting and Training Interviewers

You will need to decide how many interviewers to employ and whether they must have special qualifications. The most important selection criterion is the interviewer's ability to relate to respondents positively. An interviewer who might do a fine job of interviewing successful teachers might be totally unsuited to interview unmarried pregnant teenagers, for example.

Matching

There is evidence to indicate that matching interviewers and respondents on such variables as social class, race, age, and gender is likely to produce more valid responses.[35] The interviewer's gender is of particular concern to some researchers. Males and females traditionally have been involved mainly in superordinate-subordinate relationship patterns. These patterns can affect the interviewer-respondent relationship, especially when the interviewer is male and the respondent is female.

To clarify the influence of gender,[36] Ann Oakley identified a masculine paradigm and a feminine paradigm in interviewing. In the masculine paradigm, interviewers maintain a superordinate, emotionally neutral stance toward the respondent, and they control what the respondent says. In contrast, status differences are minimized in the feminine paradigm; the interviewers share their human side; and they give the respondents greater freedom to speak as they wish. In selecting interviewers, you might consider whether these paradigm distinctions are relevant to your study. If relevant, you will need to decide which paradigm is most appropriate to your research objectives and whether male or female interviewers are likely to be more effective. You should not assume, though, that only men can follow the masculine paradigm and only women the feminine paradigm.

Respondents as Interviewers

Some researchers recommend selecting interviewers from the respondent target population. An example of this approach is an investigation of at-risk students in urban high schools by Edwin Farrell, George Peguero, Rasheed Lindsey, and Ronald White.[37] The study involved an ethnographic perspective, and interviewing was the primary method of data collection. In designing the study, the principal investigator (Farrell) noted the difficulty of him, "a white, middle class, middle-aged academic," collecting data in "a social setting made up, for the most part, of low-income black and Hispanic adolescents."[38] Farrell dealt with the problem by recruiting students from the target population (students identified as at risk of dropping out of high school) to serve as interviewers. Seven students collected the interview data and also participated in the data analysis. Three of the students who worked for the duration of the project were listed as co-authors of the journal article reporting the study's findings.

Training of Interviewers

Once interviewers are selected, all of them should be given training. The amount of training needed will be greater as the depth of the interview increases and structure decreases. The training usually is carried out in two phases. In the first phase, the trainees study the interview guide and learn about the interview conditions (e.g., logistics, necessary controls and safeguards, topics being investigated). The researcher's hypotheses or expected results should not be discussed with the interviewers at this point, because they are likely to bias the interviewers' perceptions. Interviewers should become so familiar with the interview guide (wording, format, recording procedures, and allowable probes) that they can conduct the interview in a conversational manner without hesitating, backtracking, or needing to reread or study the guide.

In the second phase of training, trainees should conduct practice interviews and receive corrective feedback until their performance becomes polished and reaches the desired level of standardization or structure, objectivity, and reliability. Videotape recordings of practice interviews are quite effective in providing models of acceptable interviewing techniques and in giving corrective feedback. The videotape can be replayed several times so that trainees can locate procedural errors, suggest better procedures, and discuss alternative ways of dealing with problems that arise.

Depending on the interview task, some trainees may not be able to achieve the criterion standards of performance. Other trainees may not be able to stay with the project for its duration. For example, in the study of at-risk students described above, four student interviewers left the project at various points in time. If you think that these problems are likely to arise in your study, you should consider recruiting and training more interviewers than you actually need.

In the study of children with chronic illnesses, one of the research team members, two graduate students, and a professional interviewer were trained in general techniques of telephone interviewing, use of the two interview guides, and procedures for recording responses. All interviewers were checked initially by a member of the research team to ensure that they were accurate, appropriate, and consistent in their approach. Periodic checks also were made throughout the study.

Interviewers who will conduct informal or unstructured interviews typical of qualitative research require special preparation. They should have access to senior researchers who can impart their artistry and experience. Also, senior researchers can model the interviewing process and supervise new interviewers as they practice the process.

35. For examples of research studies that have explored these variables, see: Nederhof, A. J. (1981). Impact of interviewer's sex on volunteering by females. *Perceptual and Motor Skills, 52,* 25–26; Shosteck, H. (1977). Respondent militancy as a control variable for interviewer effect. *Journal of Social Issues, 33,* 36–45.
36. Oakley, A. (1981). Interviewing women: A contradiction in terms. In H. Roberts (Ed.), *Doing feminist research* (pp. 30–61). London: Routledge & Kegan.
37. Farrell, E., Peguero, G., Lindsey, R., & White, R. (1988). Giving voice to high school students: Pressure and boredom, 'ya know what I'm sayin'?' *American Educational Research Journal, 25,* 489–502.
38. Ibid., p. 490.

Step 6: Pilot-Testing the Interview

Although interviews provide valuable data, they are quite susceptible to bias. Therefore, the interview guide and procedures should be pilot-tested to ensure that they will yield reasonably unbiased data. During the pilot interviews the researcher should be alert to communication problems, evidence of inadequate motivation on the part of respondents, and other clues that suggest the need for rephrasing questions or revising the procedure. The pilot test also can be used to identify threatening questions. Norman Bradburn and his associates defined a question as threatening when 20 percent or more of the respondents feel that most people would be very uneasy talking about the topic.[39] This criterion can be employed in the pilot test to identify such questions. If there are threatening questions, procedures should be developed to lower or eliminate their threat value.

Several methods of opening the interview should be tried to determine the one that establishes the best rapport and cooperation. Also, the researcher should evaluate methods of recording interview data to determine whether adequate information is being recorded, whether the recording method causes excessive breaks in the interview situation, and whether methods for coding and analyzing the interview data are sound.

Tape recording pilot-test interviews is important even if a tape recorder will not be used during the regular interview procedure. By playing back the interview, interviewers can gain insights into their handling of the questions and become aware of problems that escaped them during the interview itself.

Interviewers also should consider selecting a subgroup from the pilot sample to check the wording of interview items. As we discussed in the section on questionnaire pilot-testing, there is evidence that the same item can be interpreted differently by different respondents. If this happens, the validity of the interview is threatened. By pretesting items, you can identify those that are ambiguous and revise them until all or most respondents interpret them similarly.

Step 7: Conducting the Interview

Researchers have discovered many interviewer behaviors that affect the quality of data yielded by the interview method. A list of recommended behaviors, compiled from various sources, is presented in Figure 8.5. Most apply to interviews conducted in the context of either quantitative research or qualitative research.

Interviewing Tasks

Researchers should consider reviewing the list of interview guidelines shown in Figure 8.5 to determine those that are important for the particular interviews that they will conduct or will train others to conduct. In addition, they should consider how they will handle the following interview tasks.[40]

Deciding how to present oneself. The interviewer will need to decide what type of personal image to present to respondents. For example, suppose that the interviewer's respondents are teachers. The interviewer might decide to present herself as both a researcher and a teacher (assuming that she has had teaching experience). In opening the interview, then, she might say something about why she is a researcher and also describe her background in teaching. The latter information might help to establish trust and rapport with respondents.

39. Bradburn, N. M., Sudman, S., et al. (1981). *Improving interview method and questionnaire design*. San Francisco: Jossey-Bass.
40. These tasks were adapted from: Fontana, A., & Frey, J. H. (1994). Interviewing: The art of science. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 361–376). Thousand Oaks, CA: Sage.

Guidelines for Conducting a Research Interview

1. Assure respondents of absolute confidentiality before beginning the interview. If necessary, explain the procedures that will be used to assure confidentiality.
2. Build rapport by engaging in small talk before beginning the interview and by using an everyday conversational style.
3. Save complex or controversial questions for the latter part of the interview after rapport has been established.
4. Explain the potential benefits of the study to the respondents.
5. The interviewer should talk less than the respondent. As a rule, the less the interviewer talks, the more information is produced.
6. Pose questions in language that is clear and meaningful to the respondent.
7. Ask questions that contain only a single idea.
8. In phrasing questions, specify the frame of reference you want the respondent to use in answering the question, for example, ask, "What do you think of the way your child's teacher handles parent-teacher conferences?" rather than "What do you think of the teacher your child has this year?" The latter question might be appropriate, however, if the goal is to determine the respondent's salient frames of reference.
9. Use simple probes when appropriate, for example, "Can you tell me more about that?"
10. Avoid contradicting or appearing to cross-examine the respondent.
11. Do not hint—either by specific comment, tone of voice, or nonverbal cues such as shaking the head—at preferred or expected responses to a particular question.
12. If a respondent seems threatened by a specific topic, move on to another one. Try returning to the topic later, with different phrasing.
13. When posing threatening or sensitive questions, ask the respondent about the behavior of friends as well as about the respondent's own behavior.
14. Do not ask many closed-form questions in succession.
15. Do not change interview topics too often.
16. Avoid leading questions, for example ask, "What is your opinion of federal aid to education?" instead of "Do you favor federal aid to education?" However, in some cases a leading question may be asked to elicit a particular type of information from the respondent.

Other aspects of the interviewer's image need to be considered as well, among them being dress, institutional affiliation, ethnicity, and life experiences. The researcher will need to consider the respondents carefully to determine which aspects of the interviewer's image are likely to be salient to them, and whether these aspects of image are likely to have an adverse or positive effect on the interview process.

Establishing rapport. The interviewer needs to decide how much rapport to establish with each respondent. Superficial rapport may be sufficient if the respondent appears comfortable with the interview process. Stronger rapport is necessary if the interviewer wishes the respondent to reveal deeply personal or sensitive information. Beyond a certain point, however, building rapport might work against the interviewer. For example, the respondent might feel so comfortable that he chooses to spend the interview time talking about matters that are irrelevant to the researcher's purposes.

Gaining trust. Trust can be an important factor in the interview process if sensitive topics are to be discussed. For example, school administrators may be quite willing to divulge their views about the best way to improve school climate. It may be an entirely different matter to ask their opinion about whether schools should provide counseling for

students who feel confused about their sexual orientation. If sensitive topics are the focus of the research study, the interviewer will need to establish a deep level of trust in order to obtain the desired data.

Understanding the respondents' language and culture. Interviewers should have a good understanding of the language and culture of their respondents, especially if nuances of language and culture are important to understanding the phenomena being investigated. For example, suppose the interviewer is collecting data from computer educators in a variety of institutional settings. In the course of an interview, a computer educator may use technical language and refer to various aspects of his workplace. If the interviewer does not understand the terminology and workplace, his ability to probe responses and take notes could be significantly impaired. This problem can be remedied to an extent if the interviewer realizes when he is not comprehending the respondent's comments and feels sufficiently comfortable to ask for clarification.

Being sensitive to nonverbal information. The interviewer will need to decide what aspects of the respondent's behavior to focus on during the interview process. Will the interviewer attend only to what the respondent says, or will she also attend to the respondent's nonverbal communication?

Raymond Gorden distinguished between four types of nonverbal communication:

> *Proxemic* communication is the use of interpersonal space to communicate attitudes, *chronemics* communication is the use of pacing of speech and length of silence in conversation, *kinesic* communication includes any body movements or postures, and *paralinguistic* communication includes all the variations in volume, pitch and quality of voice.[41]

Any one of these forms of nonverbal communication can be a significant source of research data. If desired, the interviewers can be trained to observe and take notes on their manifestations in the interview process.

Recording Interview Data

Note taking or tape recording are the usual methods for preserving the information collected in an interview. Before choosing one of these methods, the interviewer should consider carefully the advantages and disadvantages of each.

If an interview guide is used, the interviewer probably should take handwritten notes directly on a copy of the interview guide. An alternative is to use a laptop computer. As the respondent answers questions, the responses can be keyboarded directly into a computer file. The chief advantage of note taking is that it facilitates data analysis. The information is readily accessible and much of it might already have been classified into appropriate response categories by the interviewer.

A disadvantage of note taking is that it might disrupt the effectiveness of the communication between interviewer and respondent. When questions deal with simple factual information, respondents typically expect their answers to be written down, and may appear upset if they are not. On the other hand, if respondents are asked to reveal sensitive or confidential information, note taking may distract them and prevent them from giving information they otherwise might have given. In this case, the interviewer should consider delaying note taking until after the interview is completed and the respondent has left the setting. The risk is that the interviewer will forget important details, particularly those that disagree with the interviewer's expectations.

41. Gorden, R. L. (1980). *Interviewing: Strategy, techniques, and tactics* (3rd ed.). Homewood, IL: Dorsey. Quote appears on p. 335.

The use of tape recorders has several advantages over note taking for recording interview data for research. Most importantly, it reduces the tendency of interviewers to make an unconscious selection of data favoring their biases. The tape recording provides a complete verbal record, and it can be studied much more thoroughly than data in the form of interviewer notes. A tape recorder also speeds up the interview process because there is no need for extensive note taking. Furthermore, if the interview is tape-recorded, two or more individuals who are trained in your data analysis procedures can listen to the tape—or read the transcript—and code it independently. The reliability of their frequency counts or ratings can then be determined.

The main disadvantage of tape recording an interview is that the presence of the tape recorder changes the interview situation to some degree. In interviews involving highly personal information, respondents might be reluctant to express their feelings freely if they know that their responses are being recorded. The interviewer should carefully explain the purpose of the recording and gain the confidence of the respondent, so as to minimize any undesirable effects of having the interview recorded.

In doing telephone interviews, you can purchase a duplex recording jack that connects the telephone and the tape recorder that you plan to use. As soon as the phone is picked up, the recorder will begin recording, and it will record until the phone is hung up. This method of recording telephone conversations is legal as long as one of the parties on the telephone knows it is occurring. Of course, research ethics require that you inform the person to whom you are speaking that the telephone interview is being tape recorded.

You might wish to transcribe the taped material using a typewriter or word processor. If so, you can purchase a foot pedal that is connected to the tape recorder with a jack. You can listen to the tape either with earphones or though the regular speaker. When you have heard a short segment, you simply press the foot pedal to stop the tape while you record that segment. Then you press the foot pedal again to start the tape.

Software is now available to turn a computer into a tape recorder. This capability is especially useful if you have a power-book computer that you can take to sites where you plan to conduct interviews. The software allows you to record the interview, make notations in the recording, and control the replay to facilitate transcription. Other software is available to convert an audio recording or dictation into text.[42]

Whichever recording method you use, practice usually is necessary. You should reach a level of automaticity in your recording skills so that you can focus your attention on the interview process rather than on the recording process. Also, practice might identify problems and issues that are best addressed prior to formal data collection.

Step 8: Analyzing Interview Data

The analysis of responses to closed-form interview questions is straightforward. It is typical to calculate the percentage of respondents who indicated each response option for each item. For example, in the study of children with chronic illnesses, Lynch and her associates computed the percentage of school district personnel ($N = 80$) and family members ($N = 72$) who mentioned barriers to services for this type of child. The percentages are shown in Table 8.2. Note that only barriers mentioned by at least five respondents in each sample are included in the table.

42. Specific tape-recording and dictation software programs are described in: Fetterman, D. M. (1998). Webs of meaning: Computer and internet resources for educational research and instruction. *Educational Researcher, 27*(3), 22–30.

Barriers to Services for Chronically Ill Children Cited More Than Five Times by Districts or Families

Barrier	Percentage Cited	Number
District		
Lack of adequate funding	28.8	23
Lack of public and staff awareness	27.5	22
Inadequate services	12.5	10
Not enough teachers for these students	11.3	9
Children fall behind in their schoolwork	8.8	7
Children's absences	8.8	7
Uncooperative parents	8.8	7
Responsibility in the system for these students unclear	7.5	6
Family		
Teachers don't understand child's needs	9.7	7
School systems and teachers are misinformed about the illness	8.3	6

Source: Table 2 on p. 215 in: Lynch, E. W., Lewis, R. B., & Murphy, D. S. (1992). *Exceptional Children, 59,* 210–220. Copyright 1992 by The Council for Exceptional Children. Reprinted with permission.

The analysis of responses to open-form questions requires the development of a category system. An example of this approach can be found in George Kuh's study of the impact of out-of-class experiences on students in college.[43] A total of 149 college seniors at twelve institutions were interviewed by eight trained interviewers using a semistructured interview guide. The interviews were transcribed and then analyzed to determine what types of outcomes were mentioned by students. The analysis involved a five-step procedure:

1. A doctoral student read all the transcripts and developed a set of eight categories of outcomes mentioned by the college students.
2. Another individual read a sample of the transcripts, and based on her analysis, the outcomes were revised and expanded to ten categories.
3. Four readers analyzed a transcript using the set of categories developed in step 2. Their work resulted in an expanded set of 13 outcome categories.
4. The four readers analyzed four more transcripts, using the set of categories developed in step 3. Their work resulted in several minor revisions to the categories and the addition of an "other" category for miscellaneous outcomes.

43. Kuh, G. D. (1993). In their own words: What students learn outside the classroom. *American Educational Research Journal, 30,* 277–304.

5. The researcher used the category system to code all 149 transcripts. In other words, each mention of an outcome in an interview was coded as an instance of a particular category.

The following three categories illustrate the types of outcomes mentioned by students: (1) self-awareness (includes self-examination, spirituality), (2) social competence (includes capacity for intimacy, working with others, teamwork, leadership, dealing with others, assertiveness, flexibility, public speaking, communication, patience), and (3) knowledge acquisition (includes academic and course-related learning, content mastery). Kuh reported the mean number of times that each of these outcome categories was mentioned by the sample of 149 students, and the percentage of students who mentioned it 0, 1, 2, 3, 4, or 5 or more times. For example, the mean number of interview statements that were coded as the self-awareness outcome was 1.07 (standard deviation = 1.19). Forty percent of the sample did not mention it at all, 33 percent mentioned it once, 11 percent mentioned it twice, 11 percent mentioned it 3 times, 4 percent mentioned it 4 times, and 1 percent mentioned it 5 or more times.

If the interview data were collected in the context of a qualitative research study, they could be analyzed by several methods, including the grounded-theory approach described in Chapter 14. The choice of data-analysis approach will be determined in large part by the type of qualitative research that is being done. For example, an anthropologist who has done ethnographic interviews will study interview data from a different perspective than a historian who has done oral history interviews. These various perspectives are discussed in the chapters on qualitative research traditions (Chapters 15 and 16).

RECOMMENDATIONS FOR Using Questionnaires and Interviews as Research Instruments

✔ SELF-CHECK TEST

Circle the correct answer to each of the following questions.
The answers are provided at the back of the book.

1. In research, interviews differ from questionnaires in that
   a. the respondent controls the response situation.
   b. the respondent is asked to provide personal information.
   c. the question sequence and wording can vary with each respondent.
   d. responses to factual questions tend to be more accurate.

2. The most basic consideration in selecting respondents for a questionnaire study is to
   a. determine the sample size.
   b. select a sample that has the desired information.
   c. select the data-collection method that respondents prefer.
   d. study a population with which you are familiar.

3. A major problem with anonymous questionnaires, compared with questionnaires that identify the respondent, is that
   a. the return rate is much lower.
   b. respondents are less likely to provide valid information.
   c. follow-up procedures cannot be used.
   d. all of the above.

4. Pretesting in questionnaire research can be used to
   a. determine if the individuals to be sampled have sufficient knowledge to give meaningful responses.
   b. determine the likely response rate to the questionnaire.
   c. revise the questionnaire items to reduce the possibility of misinterpretations.
   d. all of the above.

5. In writing a cover letter to accompany a mailed questionnaire, a researcher would be well advised to
   a. request that the questionnaire be returned by a certain date.
   b. avoid setting a time limit for return of the questionnaire.
   c. describe the consequences if the questionnaire is not returned by a certain date.
   d. avoid associating the research project with a professional institution.

6. Research has found that respondents who do not return the first questionnaire mailed to them
   a. do not respond to follow-ups unless accompanied by a cash reward.
   b. will respond in greater numbers if more than one follow-up mailing is done.
   c. have very different personality characteristics than individuals who do complete the first questionnaire mailed to them.
   d. have poor reading skills.

7. Compared with a mailed questionnaire, the principal advantage of the interview is the
   a. low cost of data collection.
   b. depth of information collected.
   c. ease of administration.
   d. high reliability of the obtained data.

8. The research interview has the following disadvantage(s):
   a. The respondent needs a high level of verbal skills.
   b. It is not possible to probe unclear responses.
   c. It is subject to interviewer bias.
   d. All of the above.

9. The principal disadvantage of tape recording a research interview is the
   a. change that it produces in the interview situation.
   b. cost of the equipment that is required.
   c. superficiality of the obtained data.
   d. low validity of tape-recorded data compared to that of data obtained from notes.

10. It is good interview technique to
   a. ask leading questions.
   b. avoid engaging in small talk before starting the formal interview.
   c. cross-examine respondents if they seem deceptive.
   d. make sure that respondents understand the purpose of each question asked.

✦ Collecting Research Data through Observation and Content Analysis

OVERVIEW

Rather than relying solely on people's self-reports of events, many researchers prefer to make their own observations. Much of this chapter concerns the methods that quantitative and qualitative researchers use in making systematic observations of others. We also describe procedures for collecting observational data without the awareness of research participants. In the last section of the chapter, we explain how various types of artifacts and written communications found in natural settings can be analyzed to provide valuable research data.

OBJECTIVES

After studying this chapter, you should be able to

1. State the advantages and limitations of observation compared to other data-collection methods.

2. Explain the differences between descriptive, inferential, and evaluative observational variables.

3. Explain the differences between duration, frequency-count, interval, and continuous procedures for recording observations.

4. State advantages and disadvantages of using a standard observation form in a research project.

5. Explain how video recorders, audio recorders, and computers can be used to record observational data.

6. Describe an effective procedure for selecting and training observers for a quantitative research study.

7. Explain the differences between criterion-related observer reliability, intra-observer reliability, and inter-observer reliability.

8. Describe seven types of observer effects that weaken the validity and reliability of quantitative observational data, and procedures that can be used to minimize or avoid each effect.

9. State three ways in which observation in quantitative research differs from observation in qualitative research.

10. Describe the various roles that observers play in quantitative research.

11. Identify the three stages of observation in a qualitative research study.

12. Explain how observers in a qualitative research study prepare themselves and gain entry into a field setting.

13. Describe various methods for recording observational data in qualitative research.

14. Describe four types of observer effects in qualitative research and procedures that can be used to minimize each effect.

15. State the advantages and limitations of unobtrusive measures and the study of material culture.

16. Describe the steps that a quantitative researcher follows in doing a content analysis.

17. Describe the steps that a qualitative researcher follows in analyzing documents and records.

Introduction

In Chapters 7 and 8 we considered tests, questionnaires, and interviews as methods for collecting research data. All of them rely on self-report by research participants. Although self-reports usually are easy to obtain, many individuals bias the information they offer about themselves, or they cannot recall accurately the events of interest to the researcher.[1]

An alternative to self-report is to observe the behavior and social and material environment of the individuals being studied. Researchers within both quantitative and qualitative traditions have developed systematic methods for this purpose. If used properly, these observational methods avoid the inaccuracy and bias of some self-report data. For example, Lee Sechrest suggested that social attitudes like prejudice are best studied through observation in natural, real-life situations (called *naturalistic observation*) because self-reports of these attitudes often are biased by the set to give a socially desirable response.[2] Following this suggestion, some researchers have studied prejudice by observing such phenomena as the seating patterns of black and white students in college classes.[3]

Even when bias is not present in self-report data, observational methods may yield more accurate data. For example, educators have noted that teachers dominate classroom talk at the expense of student participation. But what are the actual percentages of teacher and student talk in classrooms? Self-reports by teachers or students are unlikely to yield a precise answer to this question, but an analysis of observations recorded on audiotape or videotape can do so.

Although observation is superior to self-report for some research purposes, it is more time-consuming. Individuals must be observed over a period of time to obtain reliable data, whereas tests, questionnaires, and interviews usually can yield reliable data even when the data are collected only at one point in time. Also, if the observational method is used in a quantitative study, inter-observer reliability should be established by having independent observers record data on the situation being observed. This is a difficult requirement if the researcher must rely entirely on his own resources for data collection.

This chapter has three major sections, each covering a different method of data collection: observation, nonreactive observation, and content analysis. Within each section, we treat applications of the method in quantitative and qualitative research. By organizing the chapter in this way, however, we do not mean to imply that quantitative and qualitative research differ only in how they apply what are essentially the same methods. In fact, there is a world of difference between the way that a quantitative researcher, such as a behaviorist, and a qualitative researcher, such as an ethnographer, go about making and recording observations. In general, quantitative and qualitative researchers make very dif-

1. In a review of six studies in which both observational and self-report data were collected on the same specific behaviors, none reported a clear relationship between the two types of data. See: Hook, C. M., & Rosenshine, B. V. (1974). Accuracy of teacher reports of their classroom behavior. *Review of Educational Research, 49,* 1–12.
2. Sechrest, L. (Ed.). (1979). *Unobtrusive measurement today.* San Francisco: Jossey-Bass.
3. Campbell, D. T., Kruskal, W. H., & Wallace, W. P. (1966). Seating aggregation as an index of attitude. *Sociometry, 29,* 1–15.

ferent assumptions about the nature of the social reality being observed and about their role as observers.

These assumptions are discussed in Chapter 1, which we recommend you read before studying this chapter. Then, as you read about quantitative and qualitative approaches to observation and content analysis in this chapter, we invite you to consider which approach best satisfies your own epistemological assumptions about the nature of social reality and inquiry. We also invite you to consider whether the two approaches lead to conflicting or complementary understandings of the education enterprise.

Procedures for Observation in Quantitative Research
Defining Observational Variables

In a quantitative research study, the first step in observation is to define the variables that are to be observed. To illustrate this procedure, we will refer to a study by Hirokazu Sakaguchi.[4] The purpose of his research was to determine how the English language is taught in Japanese universities. Although English is an important subject in Japanese higher education, little is known about the instructional methods used by professors. As Sakaguchi explained:

> The present study examines what actually goes on in college-level English classes, and conclusions are drawn by means of empirical observation rather than by the application of linguistic theory. In the Japanese context, this approach is somewhat unusual. Research has been done in Japan into how to make classroom teaching more effective, but it has tended to proceed on a theoretical level, and very few researchers have used practical observation and analysis as a method of research. An important reason for this is the difficulty involved in gaining access to the classroom: professors have a tendency to regard their classroom activities as sacrosanct, perhaps partly because of sensitivity to outside criticism, and in general are highly reluctant to cooperate with researchers. In this respect, this study breaks new ground.[5]

Sakaguchi selected two types of English classes for observation. One type was English literature classes taught by native Japanese instructors. The other type was English conversation classes taught by instructors whose native language is English.

Once Sakaguchi decided to use direct observation as a method, he needed to determine which aspects of classroom instruction to observe. He decided to use an adaptation of the Flanders interaction analysis system.[6] This observational system is used to collect data on ten variables. Sakaguchi used these variables, but added 15 more. Brief definitions of them are shown in Figure 9.1. More elaborate definitions, with examples of each, were used by Sakaguchi to train himself and another individual to make observations.

After deciding which aspects of classroom behavior to observe, Sakaguchi needed a method for recording the observations. Following Flanders's procedures, he decided to code each three-second interval of a classroom lesson into one of the 25 behavior categories shown in Figure 9.1. Thus, a 60-minute lesson would require 1,200 codings. (60 minutes = 3,600 seconds ÷ 3-second intervals = 1,200 intervals.)

4. Sakaguchi, H. (1993). A comparison of teaching methods in English-as-a-second-language conversation courses and reading courses in Japanese universities. *Dissertation Abstracts International, 54*(10), 3692A. (UMI No. 9405220)
5. Ibid., pp. 13–14.
6. Flanders, N. A. (1970). *Analyzing teaching behavior.* Reading, MA: Addison-Wesley.

✦ **Touchstone in Research**

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge: Cambridge University Press.

Evertson, C. M., & Green, J. L. (1986). Observation as inquiry and method. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 162–213). New York: Macmillan.

List of Observational Variables Used in Study of English Classes in Japanese Universities

Category 1 (Teacher accepts feelings). Teacher statements that reflect an awareness and unqualified acceptance of students' feelings.

Category 2 (Teacher encourages students). Praise and encouragement toward the students' questions, answers, and comments.

Category 2F (Teacher gives feedback). A quick, almost automatic response by the instructor that follows a student's statement, usually connoting approval or disapproval.

Category 3 (Teacher uses ideas of students). The instructor incorporates a student's idea into the lesson.

Category 4 (Teacher asks question). Questions asked by the teacher, except those that are directly related to the practice of English conversation.

Category 4C (Teacher asks conversational question). This applies to questions that are part of the dialogue or conversation practices.

Category 5 (Teacher lectures). This includes lecturing, expressing opinions, giving facts, interjecting thoughts, and off-hand comments.

Category 5Cr (Teacher corrects student's mistake). This consists of correcting errors in grammar, word usage, translations, pronunciation, rhythm, and intonation.

Category 5W (Teacher gives cues). Gives cues by word, expression, or sentence when a student gets stuck in the middle of an answer or translation.

Category 5C (Teacher answers conversation questions).

Category 6 (Teacher gives directions).

Category 7 (Teacher criticizes student).

Category 8 (Student responds). These are student responses to Category 4 questions.

Category 8C (Student gives conversational response). These are student responses to Category 4C questions, and to questions asked by one student to another.

Category 8R (Student engages in oral reading). This occurs mainly in literature classes when the instructor asks a student to read out loud a passage from the text.

Category 8 SR (Student engages in silent reading).

Category 8D (Student draws picture).

Category 8T (Student translates).

Category 8S (Student gives summary). This occurs mainly in literature classes when the instructor asks the student to give the main idea in Japanese from one paragraph of the English text.

Category 9 (Student talks). These are questions, answers, comments, and utterances made voluntarily—as opposed to responses to teacher questions—by the students.

Category O (Silence or confusion). Moments of non-productive confusion plus some productive silence, such as allowing students time to copy down information from the blackboard.

Category OS (Students engage in sheet work). This is time spent by students quietly working on their written exercises

Category OT (Teacher uses tape recorder). This occurs in literature classes when the instructor has students listen to tapes made by native English speakers.

Category OD (Teacher distributes handouts).

Category OB (Teacher or student writes on blackboard).

Finally, Sakaguchi needed to decide how many lessons of each instructor to observe. Limiting observation to one lesson might yield an atypical picture of the instructor's teaching style. Therefore, Sakaguchi decided to observe six lessons of each instructor in the sample. Observational data for all the instructors were collected at the same point in the university's academic year. We describe some of the findings from the analysis of these observational data below.

Types of Observational Variables

Three types of observational variables can be distinguished in quantitative research: descriptive, inferential, and evaluative. Descriptive observational variables are variables that require little inference on the part of the observer. They sometimes are called *low-inference variables* for this reason. One of their major advantages is that they generally yield reliable data. The variables shown in Figure 9.1 would be considered descriptive observational variables.

Inferential observational variables are variables that require the observer to make an inference from behavior to a construct that is presumed to underly the behavior. For example, observers might be asked to record the self-confidence with which a teacher explains a mathematical concept. Some teachers might speak with a great deal of confidence, whereas others might appear uncertain, confused, or anxious because their understanding of the topic is weak. Confidence, uncertainty, confusion, and anxiety are not behaviors but rather are psychological constructs that are inferred from behavior. For this reason they sometimes are called *high-inference variables*. It is much more difficult to collect reliable data on inferential observational variables than on descriptive observational variables.

Evaluative observational variables are variables that require not only an inference from behavior on the part of the observer but also an evaluative judgment. For example, we might be interested in obtaining ratings of the quality of the teacher's explanation of a mathematical concept. Quality is not a behavior, but rather a construct that is inferred from behavior. Also, it is a construct that is clearly evaluative in nature. Because it is difficult to make reliable observations of evaluative variables, we need to collect examples of explanations that define points along a continuum of excellent to poor explanations, and use these in training the observers.

Recording and Analyzing Observations

To ensure accurate recording, observers should be required to record data on only one observational variable at a time. For example, most observers would find it quite difficult to record various aspects of the teacher's behavior while also recording the percentage of children who are paying attention to the teacher. In this situation, the reliability of both sets of observations probably would be low. Therefore, different observers could be assigned to record each type of variable, or a single observer could alternate between recording the teacher's behavior for a specified interval and then recording the students' behavior for the next interval.

Procedures for recording observations can be classified into four major types: (1) duration, (2) frequency-count, (3) interval, and (4) continuous.

Duration recording. In **duration recording** the observer measures the elapsed time during which each target behavior occurs. A stopwatch generally is used for this purpose. It is easy to do duration recording for a single observational variable, such as the length of time a particular student is out of her seat. An observer also can record different observational variables if they do not occur at the same time. For example, the observer can record

the length of time that a particular student is on-task, off-task but not disruptive, mildly disruptive, or seriously disruptive.

Frequency-count recording. In **frequency-count recording** the observer records each time a target behavior occurs. A tally sheet typically is used for this purpose. Frequency counts are most useful in recording behaviors of short duration and behaviors whose duration is not important. For example, one of the authors (W. Borg) conducted a study in which each observer was trained to tally 13 teacher behaviors related to classroom management, such as goal-directed prompts, concurrent praise, and alerting cues.[7] The behaviors were of short duration, and no more than one behavior could occur at the same time. Interobserver reliabilities were satisfactory, ranging from .71 to .96 for the 13 behaviors.

Interval recording. Interval recording involves observing the behavior of an individual at given intervals. Sakaguchi's adaptation of Flanders's interaction analysis system is an example of this recording procedure. The instructor's or students' behavior was coded into one of the 25 observational categories every three seconds.

Once a sample of behavior has been recorded in this manner, the data must be summarized and reported to provide a meaningful description of what happened. In Sakaguchi's study, the primary research objective was to determine how the six instructors in his sample differed in their teaching style. For each observed lesson, his primary data were the number of three-second intervals in which each observational variable in Figure 9.1 occurred. For each variable, he divided the number of intervals in which it occurred by the total number of intervals that were recorded for all variables. The result of this calculation was a percentage, namely, the percentage of the total lesson time during which each variable was occurring. The final step in the analysis was to average the percentages for each variable across all six lessons that were observed for each instructor.

Table 9.1 shows a comparison of the results for two of the instructors in the sample— one a native-Japanese instructor of an English literature course, the other a native-English-speaking instructor of an English conversation course. Among the differences between the two instructors, we see that the conversation instructor spent 4.38 percent of class time encouraging students, whereas the literature instructor spent .62 percent of his time engaged in this activity. As would be expected, the students of the conversation instructor spent no class time engaged in oral reading or translating, whereas students of the literature instructor spent 19.58 percent of class time engaged in oral reading and 37.98 percent of class time engaged in translating passages from an assigned text.

The data analysis shown in Table 9.1 contributes to research knowledge about teacher and student use of instructional time in particular types of lessons. However, it does not tell us about *sequence*, that is, how the lessons unfolded over time. For example, we can surmise that each lesson started with one of the coded behaviors shown in Table 9.1, but we do not know which one. Also, we do not know what behavior was most likely to occur following, let's say, a student engagement in oral reading (code 8R).

==Various statistical techniques are available to address these questions about how observed behavior unfolds.==[8] For example, in the study by Sakaguchi, if we wished to know which behavior was most likely to occur following the occurrence of a student engaged in oral reading, we could: (1) identify all instances of the 8R code; (2) identify which code occurred immediately after the end of each 8R code; (3) count the frequency of each of these codes; and (4) divide each code frequency count by the total frequency count to yield a per-

7. Borg, W. R. (1977). Changing teacher and pupil performance with protocols. *Journal of Experimental Education, 45,* 9–18.

8. These techniques are described in Chapters 6–10 of: Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge: Cambridge University Press.

Comparison of the Observed Teaching Style of an Instructor of English Literature and an Instructor of English Conversation in a Japanese University

Code	Category	Conversation Instructor (Percentage of Lesson)	Literature Instructor (Percentage of Lesson)
1	Instructor accepts feelings	.08	.04
2	Instructor encourages	4.38	.62
2F	Instructor gives feedback	3.65	1.56
3	Instructor uses student ideas	0	0
4	Instructor asks question	13.17	7.78
4C	Instructor asks conversational question	12.25	0
5	Instructor lectures	25.33	8.30
5Cr	Instructor corrects student mistake	1.85	4.58
5W	Instructor gives cues	.92	3.44
5C	Instructor answers conversational question	1.30	0
6	Instructor gives directions	2.53	3.44
7	Instructor criticizes student	1.03	1.44
8	Student responds	1.78	2.75
8C	Student gives conversational response	22.78	0
8R	Student does oral reading	.13	19.58
8SR	Student does silent reading	0	0
8D	Student draws picture	0	0
8T	Student translates	0	37.98
8S	Student summarizes	0	0
9	Student initiates talk	.20	2.08
O	Silence or confusion	2.05	5.20
OS	Students do worksheet	6.03	1.18
OT	Instructor uses tape recorder	0	0
OD	Instructor distributes handout	.39	0
OB	Instructor or student writes on blackboard	.22	0

Source: Adapted from tables 1 and 4 on pp. 52 and 87, respectively, in: Sakaguchi, H. (1993). A comparison of teaching methods in English-as-a-second-language conversation courses and reading courses in Japanese universities. *Dissertation Abstracts International, 54*(10), 3692A. (UMI No. 9405220)

centage for each code. Using such a procedure, we might find, let's say, that after a student finished oral reading, the most frequent next behavior was for the instructor to give feedback.

Other statistical techniques for analyzing sequences of observed behavior are more complex. One of them is ==time-series analysis,== which is a statistical technique for analyzing changes in an observed variable over time. For example, suppose Sakaguchi had counted the number of English words that a particular student spoke in class over, let's say, a period of 50 class periods. Time-series analysis could be used to detect the presence of significant changes in the frequency count over this period of time.

Continuous recording. ==Continuous recording involves recording all the behavior of the target individual or individuals for a specified observation interval.== This method usually

does not focus on a specific set of observational variables. Instead, the observer typically writes a protocol, which is a chronological narrative of everything that the individual does or everything that occurs in a particular setting, such as a classroom. This method often is used in exploratory studies to help the researcher identify important behavior patterns, which subsequently are studied using one of the other methods of observational recording.

Because it is impossible to record everything in a protocol, the observer must focus on the events and contextual features that are most relevant to the research problem. To analyze the protocols, the researcher reads them, creates a content-analysis system that fits the data, and then rereads and classifies the recorded behavior into this system. This process corresponds to the steps used in quantitative content analysis of documents and other communication media, described later in the chapter.

Selecting an Observation Recording Procedure

Once you identify pertinent observational variables and their behavioral indicators, you need to select or develop a procedure for recording the observations. If a suitable procedure is not available, you might consider developing a paper-and-pencil form (sometimes called an *observation schedule*) because it is fairly easy to construct and can accommodate a variety of observational variables.

A sample observation form is shown in Figure 9.2. The form requires the observer not only to record certain behaviors as they occur, but also to evaluate some of them on a rating scale. Item 2 of the observation form in Figure 9.2 is of the latter type.

After developing a prototype of the observation form, you should try it out in a number of situations similar to those to be observed during data collection and correct any

Sample Observation Form

1. Check each question asked by the teacher into one of the following categories (observe for the first fifteen minutes of the class hour):

	Frequency	Total
a. asks student to solve a problem at blackboard	✓✓✓✓	4
b. asks student to solve a problem at his or her seat	✓✓✓✓✓✓✓	7
c. asks students if they have any questions or if they understand	✓✓	2
d. other	✓✓✓✓✓	5
	Grand Total	18

2. Each time the teacher asks a student to solve a problem, rate the problem's level of difficulty on a 5-point scale.

	Frequency	Total
1. difficult	✓✓✓	3
2.	✓	1
3. average	✓✓✓✓✓	5
4.	✓	1
5. easy	✓	1
	Grand Total	11[a]

[a]The sum here should equal the sum of categories a and b in item 1.

weaknesses you discover. For example, a common weakness of observation forms is that they require the observer to record more kinds of behavior or watch more individuals than can be done reliably. Various solutions to these problems are possible, such as employing different observers to record different behaviors or switching from the observation of one variable to another at designated intervals.

Standard Observation Forms

Instead of developing your own observation form for a research study, you may prefer to use one of the many standard observation forms that are available. These forms have several advantages. First, standard observation forms usually have reached a stage of development where they include evidence of their validity and reliability. Second, the use of a standard form saves you all the time that it would take to develop your own form. Third, most standard forms have been used in previous research studies, so you can compare your findings with theirs.

The obvious disadvantage of a standard observation form is that it may not include all the variables that you are interested in measuring. In this case you can just use part of the form that you need, and add your own procedures for assessing other variables. Keep in mind, however, that previously reported reliability and validity data may not apply if only part of an instrument is used.

Depending on your research interests, you might be able to find published collections of standard observation forms.[9] In addition, you can search the preliminary sources described in Chapter 4. To illustrate their use, we searched several electronic sources to identify standard forms for observing leadership behavior:

1. Entering the keywords *observation* and *leadership* in ERIC for the years 1984–2001 yielded 442 citations, some of which refer to observation forms.
2. Entering the keywords *observation methods* and *leadership* in PsychInfo for the years 1984–2001 yielded ten citations, some of which refer to observation forms.
3. Entering the keywords *observation* and *leadership* in ERIC's Test Locator yielded nine instruments.

Studying these citations and instruments might identify a suitable observation form. If not, the search almost certainly will provide a conceptual basis for developing your own instrument.

Use of Audiotape and Videotape Recorders

It sometimes is impractical to collect observational data while the critical behavior is occurring. One such situation is when many of the behaviors to be recorded occur at the same time or closely together. If a recording of the events is made on audiotape or videotape, it can be replayed several times for careful study and observers can count or rate the events at a convenient time. Another advantage of taping events is that it enables you to record behaviors that you did not anticipate at the outset of your study. For example, in a study of teacher praise, you might notice midway through your observations that teachers are using certain types of praise remarks that you did not anticipate in planning your observation form. If you have recorded the observations, you can replay them and

9. For example, two publications in the field of classroom research are: Borich, G. D., & Madden, S. K. (1977). *Evaluating classroom instruction: A sourcebook of instruments*. Reading, MA: Addison-Wesley Longman; Simon, A., & Boyer, E. G. (1974). *Mirrors for behavior III: An anthology of observation instruments* (3rd ed.). Philadelphia: Communication Materials Center. (See also the 1st and 2nd editions.)

reclassify the praise statements so as to include the types of praise not listed on your original observation form.

If you decide to use an audiotape recorder or videotape recorder to collect observational data, keep in mind that technical competence is required to use these devices properly. For example, you may need to develop skills in using more than one microphone and refocusing a video camera frequently.

The study of Japanese university instruction by Sakaguchi that we described above involved the use of audiotape recordings. The audiotapes could be replayed as often as necessary to ensure reliable coding of the observational variables shown in Figure 9.1.

Use of Computers and Other Electronic Devices

Various electronic devices (e.g., PalmPilots) and software programs are available for recording and analyzing observational data. The specific tasks that these devices can perform are as follows:

1. Recording and timing events and transcribing the data onto coding sheets.
2. Transferring the data from coding sheets into a computer file that then can be analyzed by a statistical package.
3. Cleaning up the data by locating coding errors and detecting "wild codes," which are codes that have no meaning in the coding system being used.
4. Interpreting the results of the data analysis. Computers have the capacity to produce a variety of graphic data representations, which can help in interpreting one's results.

If you are planning to collect observational data, you can search the literature to determine whether a suitable device has been developed and used by other researchers. For example, Ned Flanders and his colleagues have developed computer technology to record data on the types of classroom interaction variables shown in Figure 9.2.[10] The observer uses a pen-like device and special recording paper that contains a bar code for each interaction category and also for designated students, if that is desired. By moving the "pen" across the appropriate bar codes, the observer can record how long the teacher or student engaged in a particular type of behavior. This process is similar to the procedure used for charging items at the checkout counter in many stores.

After the observer has finished making observations, he inserts the "pen" in a computer interface device, which stores the observations as a computer data file. A computer program is available for analyzing this data file and creating printouts that display the results in various formats, for example, the sequence of interactions during a lesson or the amount of time that each interaction category occurred.

Selecting and Training Observers

Researchers can make their own observations, train others to make them, or share the task with others. The advantage of using other individuals is that it allows for control of the observer bias that can occur when the same individual who designs the research study and frames its hypotheses also does the observing. In addition, if two or more similarly trained observers make independent observations, you can determine the level of inter-observer reliability of the observations. Researchers have found that the most reliable observers tend to be intelligent, verbally fluent, and motivated to do a good job.[11]

10. This type of computer-assisted observation is described on pp. 133–134 in: Acheson, K. A., & Gall, M. D. (1997). *Techniques in the clinical supervision of teachers* (4th ed.). New York: Longman.
11. Harter, D. P. (Ed.). (1982). *Using observers to study behavior.* San Francisco: Jossey-Bass.

The first step in training observers is to discuss the observation form with them. Describe each item sufficiently so that they develop a thorough understanding of what is to be observed and how it is to be recorded. Also, consider making videotape recordings of situations similar to those to be observed in the study, so that you can relate actual examples of each behavior to its definition.

The next step is to set up practice observations in which all observer trainees participate. The videotapes made earlier can be used in the practice observations. Show a brief segment of the videotape, instructing trainees to record each behavior on the observation form as it occurs. Then check each trainee to determine if he correctly tallied the behaviors. If observers disagree with each other or with the criterion, replay the videotape, stopping at each behavior to discuss the most appropriate way to record it and why. During these discussions, the observer's instruction sheet should be revised to include any clarifications that arise during the training session. A few special rules typically are required to help observers make decisions about how to record unusual behavior that was not foreseen when the observation form was developed.

Determining Observer Agreement

Observational data are of no use unless they are collected by reliable observers. What does it mean for an observer to be reliable? Ted Frick and Melvyn Semmel answered this question by distinguishing between three types of observer reliability.[12] First, there is ==criterion-related observer reliability,== which is the extent to which a trained observer's scores agree with those of an expert observer, such as the researcher who developed the observation instrument. This type of reliability is important because it provides assurance that the trained observer's understanding of the variables measured by the observation instrument is the same as that of an expert. Criterion-related observer reliability typically is established by first having an expert code a videotape or audiotape of events that include all the variables measured by the observation instrument. The trained observers then code the same tape, and their data are checked for agreement with the expert's data. Criterion-related observer reliability should be checked prior to data collection, and preferably during data collection as well.

The second type of observer reliability is ==intra-observer reliability,== which is the extent to which the observer is consistent in her observational codings. This type of reliability can be established by having each observer twice code a videotape or audiotape of events similar to those that she will be asked to observe in the field. For example, observers might code a videotape on Monday and then code the same videotape a few days later. This type of reliability is not commonly established, but you should check it if possible because it provides additional assurance that your observers are reliable. Intra-observer reliability should be checked before data collection begins, and if possible during data collection.

The third type of observer reliability is ==inter-observer reliability,== which is the extent to which the observers agree with each other during actual data collection. To establish inter-observer reliability, you will need to have pairs of observers collect data on the same events. For example, suppose that you have trained five observers to collect data on sixty lessons taught by various teachers. For the sake of efficiency, you might have each observer collect data individually on ten lessons. This procedure takes care of data collection for 50 of the 60 lessons. The other ten lessons could be observed by all five observers for the purpose of determining inter-observer reliability.

12. Frick, T., & Semmel, M. I. (1978). Observer agreement and reliabilities of classroom observational measures. *Review of Educational Research, 48,* 157–184.

Frick and Semmel described various procedures for calculating the level of criterion-related observer reliability and intra-observer reliability. These procedures involve calculating a percentage of agreement or specialized correlation coefficient. The choice of procedure depends on the type of observational variable and the type of observer reliability to be determined.

Although determining observer reliability is important, it does not ensure that the final set of observational data will be reliable. Observers can agree perfectly in training or under particular field conditions, yet the typical situations that they observe may be very unstable or differ so little from each other that accurate observation to distinguish between these small differences is not possible. In other words, observer reliability is a necessary but not sufficient condition for collecting reliable observational data.

Reducing Observer Effects

An observer effect is any action by the observer that has a negative effect on the validity or reliability of the data they collect. Researchers should be aware of possible observer effects, and should take steps to avoid or minimize them. To assist researchers in this task, Carolyn Evertson and Judith Green identified various types of observer effects.[13] We describe each type below and how it can be controlled.

Effect of the observer on the observed. Unless concealed, the observer is likely to have an impact on the observed. For example, an observer entering a classroom for the first time probably will arouse the curiosity of the students and teacher. Their resulting inattentiveness may produce nonrepresentative observational data. One way to reduce this effect is for the observer to make several visits beforehand so that the students and the teacher take the intrusion for granted and behave naturally.

A more serious problem occurs when the individuals being observed are influenced by the observer's intentions. For example, suppose the purpose of the research study is to record the number and length of dyadic interactions between the teacher and students in art classes. If they learn that this is the purpose of the study, teachers are likely to increase the frequency of their dyadic interactions, particularly if they believe that this is desirable behavior. To avoid this problem, the researcher should consider informing the teachers at the outset that it is not possible to reveal the nature of the research project until it is completed, because this might affect their behavior. Also, the researcher should reassure them that the data will be kept confidential and will not reflect unfavorably on the individuals who participate.

Observer personal bias. Observer personal bias refers to errors in observational data that are traceable to characteristics of the observer. One can argue that any observations made by human beings will contain some personal bias because all of us are influenced by our experiences and beliefs. This may be true, but the design of some research studies increases the potential for observer bias to operate.[14]

Obvious sources of personal bias in observers should be looked for and eliminated if found. For example, to use an observer with a negative attitude toward ethnic minorities in a study involving observations of the creative endeavors of ethnic-minority children and other children in a nursery school clearly would be inappropriate. The observer's bias

13. Evertson, C. M., & Green, J. L. (1986). Observation as inquiry and method. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 162–213). New York: Macmillan. Evertson and Green identified ten types of observer effects. We re-organized them here into seven types.

14. Salvia, J. A., & Mersel, C. J. (1980). Observer bias: A methodological consideration in special education research. *Journal of Special Education, 14*, 261–270. The authors reviewed 153 studies having a high potential for bias and found that only 22 percent reported adequate safeguards.

almost certainly would lead to seeing more creative behavior among the other children and either ignoring, misinterpreting, or minimizing the creative efforts of ethnic-minority children in this group.

Observers have been found to produce biased data when research participants are given such labels as "emotionally disturbed" or "mentally retarded," are high or low in physical attractiveness, or have certain ethnic or socioeconomic backgrounds.[15] If such characteristics are prominent among the individuals to be observed in your study, you may be able to develop ways to minimize their influence on the observers who will be collecting data.

Rating errors. When using observational rating scales, some observers form a response set that produces errors in their ratings on these scales. A **response set** is the tendency for an observer to make a rating based on a generalized disposition about the rating task rather than on the basis of the actual behavior of the individuals. The following are three response set errors of this sort:

1. The **error of leniency** is the tendency to assign high ratings to the majority of research participants even when they differ markedly on the variable being measured.
2. The **error of central tendency** is the tendency for observers to rate all or most of the individuals whom they observe around the midpoint of the observational scales. Observers sometimes make such ratings to avoid difficult judgments.
3. The **halo effect** is the tendency for the observer's early impressions of the individual being observed to influence his ratings on all behaviors involving that individual. For example, if the observer forms an initially favorable impression of the person being observed, he may rate the individual favorably in subsequent observations.

When high-inference variables are being observed, the magnitude of these observer rating errors can be so large that the resulting ratings are virtually meaningless. For example, a study of the ratings of student teachers by cooperating teachers revealed so much halo effect and error of leniency that the validity of the entire rating system was called into question.[16] For example, the mean "attitude" rating of the 161 student teachers who were observed was 4.85 out of a possible 5.00, which suggests a strong error of leniency. If such response sets occur, you need to either reconceptualize the rating scale or select and train observers more carefully.

Observer contamination. Observer contamination occurs when the observer's knowledge of certain data in a study influences the data that he records about other variables. For example, suppose that we are doing a study of the human relations skills of successful elementary school principals. Unsuccessful and successful principals could be identified by a composite of nominations made by teachers, parents, and school superintendents. Observers then are trained to observe the performance of the successful and unsuccessful principals in faculty meetings and evaluate them on certain human relations skills. If the observers know beforehand which principals have been classified as successful and which as unsuccessful, they almost certainly will be influenced by this knowledge when they collect observational data about the principals' behavior. The obvious solution to the problem is to keep possibly contaminating information from the observers.

15. Ibid.
16. Phelps, L., Schmitz, C. D., & Boatright, B. (1986). The effects of halo and leniency on cooperating teacher reports using Likert-type rating scales. *Journal of Educational Research, 79*, 151–154.

Observer omissions. An **observer omission** is the failure to record the occurrence of a behavior that fits one of the categories on the observational schedule. This failure can have several causes. One of them is personal bias, which we discussed above. Because of personal bias, the observer may overlook the occurrence of desirable behavior by an individual toward whom he has a negative bias. Another possibility is that the behaviors to be observed occur simultaneously or so rapidly that the observer is unable to record all of them. The opposite situation also is possible: The behavior to be observed occurs so infrequently that the observer fails to notice it.

Observation errors due to omissions can be detected during the development of the observation form or during observer training. You may find it necessary to simplify the observation schedule or to assign multiple observers to a setting, with each observer responsible for recording data on different observational variables. In the case of infrequently occurring variables, you may need to provide cues and reminders to maintain observers' vigilance.

Observer drift. Once observers have been trained to the desired level of agreement and accuracy, they should start collecting data promptly because a delay will result in some loss of observer skills. Also, if the observations are to extend for more than one week, you should hold a weekly refresher training session for all observers. If this is not done, the observational data will become less reliable because of **observer drift**, which is the tendency for observers gradually to redefine the observational variables, so that the data that they collect no longer reflect the definitions that they learned during training.

Reliability decay. Research evidence suggests that observers should be checked frequently during the course of the study to keep them performing at a satisfactory level. Otherwise, the observational data are subject to **reliability decay**, which is the tendency for observational data recorded during the later phases of data collection to be less reliable than those collected earlier.

Paul Taplin and John Reid compared "decay" in reliability for three groups of observers: (1) those who were told they would not be checked, (2) those told they would be spot-checked at regular intervals, and (3) those told they would be checked on a random basis.[17] The randomly checked group maintained the highest level of reliability, followed by the spot-checked group; the not-checked group had the lowest reliability. The spot-checked group performed very well in the sessions when they knew they would be checked, and very poorly in the sessions when they thought they would not be checked. Therefore, the problem seems to be one of motivation, so you should do whatever possible to maintain observers' motivation. For example, you can try to convince observers of the importance of their task, schedule sessions so as to avoid observer fatigue, inform them that you will check their performance on a random basis, carry out frequent random checks, and give them frequent feedback on their reliability.

Procedures for Observation in Qualitative Research

Observation in qualitative research differs from observation in quantitative research in three ways. One difference is that observers in a qualitative study do not seek to remain neutral or "objective" about the phenomena being observed. They may include their own feelings and experiences in interpreting their observations. An example is Cathy Evans and Donna Eder's study of social isolation, which included in-depth observations of informal

✦ *Touchstone in Research*

Angrosino, M. V., & Mays de Pérez, K. A. (2000). Rethinking observation: From method to context. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 673–702). Thousand Oaks, CA: Sage.

17. Taplin, P. S., & Reid, J. B. (1973). Effects of instructional set and experimenter influence on observer reliability. *Child Development, 44,* 547–554.

activities among middle school students during lunch.[18] The lunchroom observations included taking notes on the behavior of students who sat together while eating lunch toward "isolates"—students who spent most of their lunch period alone.

The researchers observed many incidents involving ridicule of isolates, and all four observers "reported that witnessing such events was a source of emotional distress for them."[19] After one incident of this type, Evans wrote these field notes, which were included in the published research report:

> I was utterly and completely disgusted. I guess part of my problem was that I was disillusioned too—I thought Janice and Patty were above it. I didn't initiate conversation with Jenny because it would have targeted her more and I didn't reprimand the girls in my group because it would have possibly been more embarrassing to Jenny and I might as well witness what there is to witness, even if it's grossly unpleasant.[20]

Thus, these qualitative researchers sought to avoid criticizing or taking sides with specific participants, but still clearly considered their own reactions to events to be a legitimate part of the study, and worthy of reporting.

The second difference between quantitative and qualitative observation is that the focus of qualitative observation is much more emergent. In contrast, data collection in quantitative research generally is driven by a priori hypotheses, questions, or objectives. At any point in the process, qualitative observers are free to shift their attention to new phenomena as new research questions emerge.

The third difference is that the focus of observation generally is much wider in qualitative research. In quantitative research, observers tend to concentrate on specific aspects of behavior and to ignore context. In qualitative research, however, observers look at behavior and its environmental setting from a holistic perspective.

The Purpose of Observation in Qualitative Research

Two common methods of data collection in qualitative research—interviews and analysis of documents—involve words uttered or written by the participants in the natural setting. This information is limited by participants' knowledge, memory, and ability to convey information clearly and accurately and, also, by how they wish to be perceived by outsiders such as the researchers. Observation, in contrast, allows researchers to formulate their own version of what is occurring, independent of the participants. The inclusion of selected observations in a researcher's report provides a more complete description of phenomena than would be possible by just referring to interview statements or documents. Just as important, observations provide an alternate source of data for verifying the information obtained by other methods. Their use for this purpose is called *triangulation,* which is explained in Chapter 14.

An ethnographic study conducted by Katherine Rosier and William Corsaro provides an example of observation as one of several data sources.[21] Their study tested the validity of the common stereotype that the educational and economic problems of many African-American youth stem from deficiencies of the families in which they are raised. One aspect of the study involved interviewing parents and observing in the homes of children enrolled

18. Evans, C., & Eder, D. (1993). "NO EXIT": Processes of social isolation in the middle school. *Journal of Contemporary Ethnography, 22,* 139–170.
19. Ibid., p. 146.
20. Ibid., p. 146.
21. Rosier, K. B., & Corsaro, W. A. (1993). Competent parents, complex lives: Managing parenthood in poverty. *Journal of Contemporary Ethnography, 22,* 171–204.

in Head Start. When visiting the homes, the researchers used direct observation to check the validity of parents' claims that they regularly engaged their children in structured learning activities at home and provided educational toys and supplies despite their limited budget. Considerable observational evidence was found to support such claims, including finished projects displayed on refrigerators and walls; tables cluttered with writing instruments, paper, and works in progress; and educational toys and books. Parents also presented the researchers with samples of their children's schoolwork.

The researchers described a day that they spent observing Cymira, one of the Head Start children, and four other children who were being cared for by Cymira's mother, Rhonda. While the children were playing outside, Rhonda got out a variety of both purchased and handmade educational materials, suited to varying ages. She then called all five children into the house. During the hour when they stayed indoors, Rhonda "checked and encouraged the children's work, and the older children assisted the younger ones."[22] Their careful, on-site observation helped the researchers confirm the validity of the finding from parent interviews that most parents of poor African American children seek to instill educational skills and values in their children at an early age.

In Chapter 15 we explain how qualitative research reflects various traditions, such as ethnography, cultural studies, and cognitive psychology. The purpose and form of observation varies across these traditions. In the following sections, we describe general methods that are used in collecting qualitative observational data. If your proposed study follows a particular tradition of qualitative research, you should study the specific ways in which researchers working within that tradition have used observation as a data-collection method.

Identifying the Observers and Their Roles

The observer role in qualitative research varies along a continuum from complete observer to complete participant. At the extreme role of **complete observer,** the researcher maintains a posture of detachment from the setting being studied. At the extreme role of **complete participant,** the researcher studies a setting in which she already is a member or becomes converted to genuine membership during the course of the research. For example, David Hayano reported on his observations of fellow denizens of California's all-night card rooms.[23]

Between these two extremes are the observer-participant and participant-observer roles. In the **observer-participant role,** the researcher acts primarily as an observer, entering the setting only to gather data and interacting only casually and nondirectly with individuals or groups while engaged in observation. In the **participant-observer role,** the researcher observes and interacts closely enough with individuals to establish a meaningful identity within their group; however, the researcher does not engage in activities that are at the core of the group's identity. For example, Peter Adler, a sociologist and professor, played a participant-observer role in a study of college athletes at his university.[24] He maintained his researcher identity while actively participating with the basketball team and coaching staff in various roles, including insider, expert, and celebrity—but not in the core roles of coach, trainer, or player.

The researcher who designs a qualitative study most likely will be the observer during the data-collection phase. Another option is for a team, which may include the researcher,

22. Ibid., p. 182.
23. Hayano, D. (1982). *Poker faces.* Berkeley, CA: University of California Press.
24. Adler, P. A., & Adler, P. (1991). *Backboards and blackboards.* New York: Columbia University Press.

to make the observations. Use of multiple observers lessens the burden on each observer and allows for more observation time overall. Furthermore, if the observers are diverse in factors relevant to the phenomena being studied (e.g., gender and age), they can enhance the validity of observations by cross-checking each others' findings and eliminating inaccurate interpretations. In the study of peer treatment of social isolates described above, two other young adults served as lunchroom observers along with the two researchers.

Preparing for Observation

An observer in a qualitative research project can prepare by serving as an apprentice to an expert in the type of observation being planned. A period of apprenticeship is desirable because qualitative observation skills are complex and subtle; they do not lend themselves to the type of training procedures for quantitative observation that we described earlier in the chapter. By working alongside an expert, a novice observer gradually can develop an understanding of how to focus her observations and to shift the focus across the three stages (descriptive, focused, and selected) described below.

Although not a substitute for apprenticeship, some university programs provide courses in which you can develop particular skills needed in qualitative observation. These skills include the ability to write descriptions of observed events that are objective rather than interpretive, to incorporate rich detail about observed events into field notes, and to convert rough, handwritten field notes into polished typed reports. Training in the use of videotape and audiotape recording equipment and in field work techniques may be provided as well.

Once a qualitative research study is underway, you might encounter problems and issues unique to your field setting. It is helpful in such situations to maintain contact with an expert who can advise you. This consultation process does not violate the integrity of the study because, in qualitative research, the observation methodology is free to change as the researcher develops new insights into the phenomena being studied.

Determining the Focus of Observation

The focus of a qualitative researcher's observations is likely to shift from early to later stages of a study. According to James Spradley, this process of shifting typically includes three stages.[25] First is the **descriptive stage,** when observations tend to be unfocused and general in scope, providing a base from which the observers can branch out in many directions. Second is the **focused stage,** when the observers have identified features of the phenomena under study that are of greatest interest and begin to direct their attention to collecting deeper information about this narrower range of features. Finally, there is the **selected stage,** when research questions or problems emerge, and the observers' focus shifts to refining and deepening their understanding of the specific elements that have emerged as theoretically or empirically most essential. Observational data are gathered until researchers achieve **theoretical saturation,** that is, when newly gathered findings essentially replicate earlier ones.[26]

The question of research focus also involves the decision as to what to observe at any given moment, and how to ensure that everything of potential interest is attended to. With respect to ensuring thorough coverage of the phenomena of interest, Norman Denzin suggested that all observational field notes should contain explicit reference to the following

25. Spradley, J. P. (1980). *Participant observation.* Fort Worth, TX: Harcourt.
26. Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

elements: participants, interactions, routines, rituals, temporal elements, interpretations, and social organization of the participants.[27] Sharan Merriam presented a similar list, but also included the setting (i.e., the physical environment, the context, and the kinds of behavior that the setting encourages, permits, discourages, or prevents) as well as subtle factors.[28] Subtle factors include informal and unplanned activities, symbolic and connotative meanings of words, nonverbal communication such as dress and physical space, and what does *not* happen, especially if it ought to have happened.

A study by Robert Prophet and Patricia Rowell illustrates the features of a phenomenon on which observers might focus, and how the focus of observation shifts as the research project progresses.[29] Prophet and Rowell investigated teacher-student interaction in science classes in the African country of Botswana. The classes were at the junior secondary level and were taught in the English language. They described the starting point for the study as follows:

> [We had] a desire to gain information and insights into the realities of life in science classrooms in the junior secondary schools. This goal was pursued using a conceptual framework in which teaching style and the use of knowledge exemplify power relationships in the classroom.[30]

Their descriptive stage of observation involved documenting classroom interactions by sitting discreetly at the back of the classroom and writing extensive notes. Excerpts from these field notes were included in the research report. The notes primarily documented utterances by the teacher, by individual students, and by students speaking in unison. However, the observers also noted specific actions, such as when students raised their hands in order to be called on, when a student's comment was ignored by the teacher, or when the teacher wrote something on the board.

Prophet and Rowell stated what their initial observations and interviews revealed:

> ... the teachers generate learning experiences that are teacher-centered and which place the students in the role of passive recipients of knowledge. ... The actual pedagogical principles underpinning teaching styles ... appear to be based on rote learning. ... These practices are so widespread and so taken for granted by all the participants that it is tempting to say simply, "This is what happens in schools."[31]

In their report, Prophet and Rowell expressed their growing sense of a major disparity between stated curriculum aims (i.e., that the teaching situation should be student-centered) and the actual authoritarian, teacher-centered classroom environment that they consistently observed.

Through their concurrent work at the University of Botswana, Prophet and Rowell realized that preservice teachers specializing in science education found "hands-on" science (e.g., laboratory experimentation) arduous and placed heavy reliance on rote learning in their own educational endeavors. Because they had seen similar problems among secondary school students in their research study, Prophet and Rowell speculated that, "the seeds for future difficulties were being sown in the teacher education program."[32]

27. Denzin, N. K. (1989). *The research act* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
28. Merriam, S. B. (1998). *Qualitative research and case study applications* (revised & expanded). San Francisco: Jossey-Bass.
29. Prophet, R. B., & Rowell, P. M. (1993). Coping and control: Science teaching strategies in Botswana. *International Journal of Qualitative Studies in Education, 6,* 197–209.
30. Ibid., p. 198.
31. Ibid., p. 201.
32. Ibid., p. 205.

During the final, selected-observation stage of their study, Prophet and Rowell paid particular attention to classroom activities designed to develop students' manipulative skills. Further observations, one of which is described in the research report, indicated that

> ... the teaching of manipulative skills followed the same teacher-centered routine as described earlier, with students often not participating and teachers presenting instructions and information in an unclear verbal style. Very little student-based practical work was observed. ...[33]

This teaching style was widely discrepant from the syllabus for the preservice teacher education program in science education.

As another part of the selected stage of their research, Prophet and Rowell developed a test of science skills and administered it to a small sample of students at each school. Students in general were found to perform poorly on the manipulative skills that were tested. In the conclusion of their research report, Prophet and Rowell summarized and interpreted their findings in relation to their previously stated formulations concerning power relationships in classrooms.

We see in this research study how narrowing and deepening the focus of field observations can foster the emergence of qualitative interpretations, which constitute the primary findings of the study.

Gaining Entry into the Field Setting

As one would expect from the emergent nature of qualitative research, there are no strict rules about how to enter a field setting to make observations. You will need to develop a procedure based on the characteristics of the field setting and its members and on where you intend to situate yourself along the continuum of complete participant to complete observer.

You can gain insights into developing your procedure by consulting with expert qualitative researchers and by reading reports of their studies. For example, the study of social isolates in middle schools by Evans and Eders, described earlier in the chapter, is instructive. They describe their procedure for gaining entry into school lunchrooms as follows:

> Our strategy was to enter the lunchroom setting as peers rather than as authority figures. We wore jeans and other casual clothing as did most of the students. We made a point of never affiliating with teachers or other adults in this setting. This meant that some of us were sometimes included when teachers chastised our groups for initiating food fights and other disruptive activities! We initially entered the lunchroom environment as we imagined a new student in school would do, by sitting alone or next to a single person and starting conversations with students sitting nearby. ... although we informed everyone that we were from the local university and that we were studying adolescent friendship, we did not take notes openly ... We did not inform the students of our ages, and although the age difference between the students and the researchers varied between 10 and 20 years, the students' consistent comments indicated they generally viewed us as ... undergraduate students in our late teens. ... The researchers' typical stance with groups to which we belonged was to become "a quiet member" of the group. We adopted a listening, receptive, nonjudgmental attitude toward other members. ... When we did not report them for swearing or other school rule violations, their trust toward us increased dramatically and they assured other students that we were "okay."[34]

33. Ibid., p. 205.
34. Evans & Eder, Processes of social isolation, pp. 145–146.

This description shows the researchers' insights about the need to present themselves to the adolescents under investigation as nonthreatening and friendly. If the same researchers were to shift their research focus to teachers or administrators in the same school, one would expect them to adopt a more professional and mature appearance and manner, but to maintain the same level of neutrality and discretion with those groups as they had with the students, thereby demonstrating their worthiness of trust.

Recording Observations

Qualitative researchers can make use of the same range of methods used by quantitative researchers to make a permanent record of their observations. For example, they can take written field notes using a lap-top computer. Another option is to dictate notes into an audiotape recorder. A stenomask can be useful for this purpose.[35] A stenomask is a sound-shielded microphone attached to a portable tape recorder that is worn on a shoulder strap. The observer can speak into the microphone while an activity is occurring without people nearby being able to hear the dictation.

In some field settings, an observer taking notes on a tablet or in a notebook might distract participants or cause the observer to miss important aspects of events. It may be possible to take a few notes surreptitiously in such situations. Qualitative researchers even have been known to write field notes on toilet paper or inside matchbook covers in order to conceal their role as observers!

If you cannot take notes while in the field setting, you will need to remember what occurred and arrange to make notes soon after leaving the scene. Even if you cannot make complete field notes right away, you should at least try to write a summary of the sequence of events and noteworthy statements. You can use this summary later to stimulate your writing of a more extensive set of notes.

The following are features of good field notes in a qualitative research study.

Field notes should be descriptive and reflective. **Descriptive information** includes verbal portraits of the research participants, reconstruction of dialogue, description of the physical setting, accounts of particular events, and descriptions of the observer's behavior. **Reflective information** includes the researcher's personal account of the course of inquiry, and may contain the following elements: reflections on the methods of data collection and analysis, reflections on ethical dilemmas and conflict, reflections on the observer's frame of mind, and emerging interpretations. An example of field notes that include both types of information is shown in Figure 9.3. The notes that are reflective in nature are in italics.

Field notes should be detailed and concrete. Observers should strive to write field notes that are detailed and concrete, not vague and overgeneralized. Michael Patton provided the following examples of field notes that are at these two extremes:[36]

Vague and Overgeneralized Field Notes	**Detailed and Concrete Field Notes**
The client was quite hostile toward the staff person.	When Judy, the senior staff member, told her that she could not do what she wanted to do, the client began to yell at Judy, telling her that she couldn't control her

35. Stenomasks are described on pp. 248–249 in: Patton, M. (1990). *Qualitative evaluation and research methods* (2nd ed.). Thousand Oaks, CA: Sage.
36. Patton, *Qualitative evaluation*, p. 240.

Example of Field Notes with Observer Comments

> B turned on the overhead projector and proceeded to show some examples. The first one was a colorful slide of the planets.
>
> *OC—B must like planets. There is a model of the solar system, in similar colors in B's office.*
>
> B pointed out that this particular slide would be useful in having students identify the planets and that the slide would not be very helpful in explaining planetary motion. The students were also shown an example of masking. (The names of the planets were masked by pieces of cardboard so that they could be revealed by the teacher, as required.)
>
> *OC—The students were attentive, but quiet, too quiet I thought. What's happening here is that the teacher is not asking enough questions. For example, B could have asked the students what the flaps were for, and why would you want to do such a thing. Instead of telling what the slide was good for, asking what it would be good for. Nice locus of control issue.*
>
> The second transparency's subject was the water cycle, a slide consisting of the main and two overlays. B explained how an overhead of this kind could be used to describe a process.
>
> *OC—The students were quiet, no questions.*

Source: Pages 100–101 in: Brandt, R. *Observation of B's lesson on overhead transparencies.* Unpublished report, University of Georgia, Apr. 30, 1987. Reprinted in Merriam, S. B. (1988). *Case study research in education: A qualitative approach.* San Francisco: Jossey-Bass.
Note: Italicized field notes are reflective comments by the observer. OC is an abbreviation for observer comments.

life, that she was on nothing but a "power trip," that she'd "like to beat the shit out of her," and that she could just "go to hell." She shook her fist in Judy's face and stomped out of the room, leaving Judy standing there with her mouth open, looking amazed.

The excerpt on the right obviously is much more specific and descriptive than that on the left. Also, the language reflects low-inference observation, which is more helpful than high-inference observation when the researcher is ready to look for themes and patterns in the field notes and other data sources.

Field notes should include visual details when appropriate. Field notes need not be limited to words. For example, an observer might draw a sketch of the layout of the physical setting in which the observed activities are occurring. If visual details are worthy of even more attention, the researcher can create a documentary-style visual record by making videotape recordings or taking photographs. Anthropologists and sociologists frequently use such data-collection methods in their fieldwork.[37] At their best, visual records illuminate important aspects of culture and social interaction. However, some critics have argued that documentary-style visual records, particularly those of early-20th-century anthropologists, while seemingly neutral, "reified the relationships of superiority and inferiority endemic of colonialism."[38] This criticism suggests the need for qualitative researchers

37. Harper, D. (2000). Reimagining visual methods: Galileo to Neuromancer. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 717–732). Thousand Oaks, CA: Sage.
38. Ibid., p. 728.

to analyze their motives and stance in using visual recording devices to select the particular scenes and objects for preservation as recorded images.

A case study by Terry Wood, Paul Cobb, and Erna Yackel illustrates the use of videotape recordings in qualitative research.[39] The researchers examined how one elementary school teacher changed her approach to teaching mathematics while participating in an ongoing research project that involved constructivist views of learning. Every mathematics lesson taught by the teacher for an entire school year was videotaped, and two cameras were used to focus on four selected pairs of children during small-group activities. The videotape recordings, supplemented by field notes and copies of the children's classwork, were the main source of data used to document changes in students' mathematical understanding and social interaction in small groups.

Dealing with Observer Effects

Quantitative research operates on the premise that observations should be independent of the particular individual who is making them. Thus, an effort is made to minimize observer bias and to control for possible effects of the observer on what is observed. Qualitative researchers take a different stance toward their influence on the phenomena being observed. As observers, they consider their biases and personal reactions to be part of the "scene" that is observed. Therefore, qualitative researchers do not use the positivist criterion of objectivity in deciding whether their observations are of high quality. Instead, they use the procedures described below to deal with validity issues that arise in qualitative observation.[40]

Reactions of program participants and staff to the observer's presence. The recommended approach is that qualitative researchers should strive not to overestimate or underestimate their effects on what is observed, but they should describe and analyze these effects as part of the research project.

Effects on the observer during the course of the study. The approach recommended for dealing with this issue is similar to that suggested for the previous issue, namely, to realize that the observer will have reactions and to record them. The excerpt from Evans and Eder's field notes, in which the observer reported her distress at witnessing the mistreatment of social isolates by their middle school peers, typifies qualitative observers' attention to the effects of observations on themselves.

The observer's personal predispositions or biases. To address this issue, the researcher should use established procedures for validating and verifying data analyses, so as to reduce any distortions that may have been introduced by the researcher's predispositions. These procedures include active efforts to test rival explanations for research findings, the use of both qualitative and quantitative research methods to examine a phenomenon, the use of multiple observers and researchers, the examination of findings from various theoretical perspectives, and a reporting of the research project in sufficient detail that readers can "audit" the findings.

Observer incompetence. Qualitative observation data will be useless, and even misleading, if the observers have had insufficient preparation to do the type of data collection required by the research problem and approach. The solution to this problem is obvious. Observers should be thoroughly trained and otherwise prepared prior to collecting data in the field.

39. Wood, T., Cobb, P., & Yackel, E. (1991). Change in teaching mathematics: A case study. *American Educational Research Journal, 28,* 587–616.
40. These issues and procedures also are discussed in Chapter 14.

Analyzing Qualitative Observational Data

When the fieldwork phase of a qualitative research study is completed, the researchers are likely to have an extensive set of field notes and visual data that serve as a record of their observations. All these data need to be analyzed, interpreted, and reported. We describe each of these processes in the chapter on the case study method (Chapter 14). You will find that the procedures for analyzing, interpreting, and reporting observational data are essentially the same as those for other types of qualitative data, such as interview notes and documents found in the field setting.

Nonreactive Observation

Unobtrusive Measures in Quantitative Research

We explained above how observer effects can weaken the validity and reliability of observational data. For example, the presence of an observer can affect the behavior of the observed individuals such that it becomes atypical. This and other observer effects can be avoided through the use of unobtrusive measures—sometimes called *nonreactive measures.* Unobtrusive measures are characterized by the fact that the data are collected in a natural setting, and the individuals are unaware that they are being observed.

Suppose we are interested in students' use of computer technology in their studies. It would be difficult to observe students throughout the day, not knowing if and when they might use a computer. Unobtrusive measurement can help us with this problem. We could think of situations in which students might use a computer, and whether a product or "residue" would be produced that we could examine later. An obvious possibility is the use of computers for word processing of school assignments. We could ask the students' teachers for permission to examine the students' assignments so that we can observe which ones appear to be word-processed documents and which do not.

In this approach the students are unaware that they are being observed, but their teachers are aware. Perhaps we can think of situations involving computer use that are unobtrusive for both groups. Suppose the schools that we are observing make computers available for students in a technology center and in the school's library. We can obtain unobtrusive software that can be installed in these computers for the purpose of recording their duration of use and even the users' keystrokes. Of course, this approach raises ethical questions, which we consider below.

Examples of unobtrusive measures that have been used in research studies are listed in Figure 9.4. Measures of this type are especially useful when used in conjunction with conventional reactive measures because they involve such a different approach to measurement. If we use several different kinds of instruments to measure the same variables, and they yield similar results, we can be much more confident that our results are valid.

Potential Limitations of Unobtrusive Measures

Validity. The validity of some unobtrusive measures is uncertain. For example, the lost-letter technique has been used as an unobtrusive measure in research studies.[41] This technique involves making a large number of copies of various bogus letters, each of which is addressed to an organization that reflects a different opinion on an issue. The assumption is that an individual who finds a letter is more likely to mail it if the address represents an opinion that

41. For an interesting application of this technique, see: Farrington, D. P., & Knight, B. J. (1980). Stealing from a lost letter: Effects of victim characteristics. *Criminal Justice and Behavior, 7,* 423–435.

Examples of Unobtrusive Measures in Research Studies

1. The floor tiles around the hatching-chick exhibit at Chicago's Museum of Science and Industry must be replaced every six weeks. Tiles in other parts of the museum need not be replaced for years. The selective erosion of tiles, indexed by the replacement rate, is a measure of relative popularity of exhibits.
2. One investigator wanted to learn the level of whiskey consumption in a town that was officially "dry." He did so by counting empty bottles in trash cans.
3. The degree of fear induced by a ghost-story-telling session can be measured by noting the shrinking diameter of a circle of seated children.
4. Chinese jade dealers have used the pupil dilation of their customers as a measure of the client's interest in particular stones. In 1872 Darwin noted this same variable as an index of fear.
5. Library withdrawals were used to demonstrate the effect of the introduction of television into a community. Fiction titles dropped, while nonfiction titles were unaffected.
6. The influence of the rate of interaction in managerial recruitment is shown by the overrepresentation of baseball managers who were infielders or catchers (high interaction positions) during their playing days.
7. Sir Francis Galton employed surveying hardware to estimate the bodily dimensions of African women whose language he did not speak.
8. Children's level of interest in Christmas was demonstrated by distortions in the size of their Santa Claus drawings.
9. Racial attitudes in two colleges were compared by noting the degree of clustering of African Americans and whites in lecture halls.

Source: Adapted from Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences* (2nd ed.). Boston: Houghton Mifflin.

he or she supports. If this is true, the rate of return for letters representing different opinions will reflect the percentage of persons holding each opinion in the community under investigation. In several election studies, however, the proportion of letters returned failed to predict the election results, thus raising doubts about the validity of the technique.

Reliability. The reliability of unobtrusive measures often is difficult to establish. Even for unobtrusive measures for which reliability can be computed, the data are of limited use to other researchers because most such measures are designed to study a very specific attitude or behavior pattern and are rarely used more than once. In contrast, conventional measures such as achievement tests, personality inventories, and attitude scales are used in many studies. Over a period of time, a useful body of knowledge about such measures is developed, and it is easier to obtain estimates of their reliability.

Ethical Considerations. The use of unobtrusive measures raises ethical issues involving informed consent and invasion of privacy. (These issues are discussed in Chapter 3.) Collection of nonreactive data in public settings usually is not regarded as an invasion of privacy. However, data collection in public areas where individuals would expect their behavior to be private, such as public restrooms, has been challenged as an invasion of privacy. Spying on an individual's private behavior (e.g., placing listening devices in the individual's home or office) clearly is an invasion of privacy.

Informed consent poses a difficult problem in unobtrusive measurement because its main purpose is to collect data without the awareness of the individuals being studied. The act of requesting informed consent makes individuals aware that they will be studied

and thus jeopardizes the nonreactivity of any subsequent measurement. Therefore, it is essential in some situations to forego informed consent.

An institutional review board (see Chapter 3) may look favorably on a request to forego informed consent if the the researcher can demonstrate the following conditions: The individuals being studied will incur no risk, the anonymity of the participants will be maintained, it is impossible to conduct the study under the condition of informed consent, and the study promises to produce significant benefits.

Qualitative Observation of Material Culture

The same material objects and residues of human actions that are studied by quantitative researchers can be phenomena of interest to qualitative researchers. Whereas quantitative researchers analyze these phenomena using narrowly focused, numerical measures (typically, frequency counts), qualitative researchers have a holistic perspective and examine these phenomena as expressions of how individuals and groups construct their social reality.

Ian Hodder used the term material culture to refer to the various objects created by different groups throughout history.[42] (Some researchers refer to them as *artifacts.*) He distinguished between two types of material culture. First is material culture that serves a communicative, representational function. Text and signs are examples of this type of material culture. We discuss the study of text and signs by qualitative researchers later in the chapter.

The second type of material culture, which we call **practice-oriented material culture** is distinguished by its association with particular practices and meanings. Tools are a good example of practice-oriented material culture. A hammer, for instance, does not have a communicative function. Rather, it is associated with particular uses, especially within the construction industry. Also, a hammer has different meanings for different individuals. One person may associate a hammer with injury, whereas another individual may have positive associations—work for pay, building useful products. Still others might form symbolic associations; for example, the song, "If I Had a Hammer," may come to mind.

Objects other than tools also can be manifestations of practice-oriented material culture. Lilies, for example, often are used for decorative purposes in our culture. They also are associated with funerals in the minds of some people.

Observation of practice-oriented material culture is important because it gives qualitative researchers access to those whose voices are silent, for example, groups whose experiences are not recorded in written form by the dominant culture. Another example is dead societies, especially those that left no written record, but only practice-oriented objects and material residues of their use.

Despite its importance, the use of practice-oriented material culture as a data source is problematic because qualitative researchers usually cannot check the validity of their interpretations of the social reality that it represents by interviewing the individuals who produced it. Archaeologists have taken the lead in developing methods to deal with this problem because they often find themselves in the situation of interpreting objects and fragments of objects from ancient cultures.[43]

Practice-oriented material culture is worthy of observation by educational researchers. Many specialized tools have been developed or adopted by educators over

42. Hodder, I. (2000). The interpretation of documents and material culture. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 703–715). Thousand Oaks, CA: Sage.
43. Ibid.

time—among them, chalkboards, desks, classrooms, bulletin boards, computers, and video recorders. Students, too, develop or purchase certain tools, such as notebooks, binders, pencils, staplers, calculators, and erasers. The study of these tools may provide valuable insights into how these groups construct the social context within which they cope with learning tasks and deal with other aspects of their public and private lives.

Content Analysis of Documents and Other Communication Media

Thus far in this chapter, our focus has been on methods for observing human behavior and features of the environment in which the behavior occurs. An important feature of human environments is the messages that people encode in various forms:

Written documents: Written materials include textbooks, students' completed home-work assignments, tests, computer printouts of school data, newspapers, and memoranda.
Visual media: Photographs, posters, and drawings are examples of visual materials that may be analyzed.
Audio media: The researcher can analyze audiotape recordings, laser disc recordings, or radio programs.
Combinations of media: It may be desirable to analyze a variety of types of media, such as TV programs and CD-ROM discs that combine print, visual images, and sound.

The content of these materials comprises messages from one individual or group to another individual or group. These messages are the object of study in some research projects. Textbooks are of particular interest because they convey much about the school curriculum. Textbooks and other written materials usually are called *documents* by researchers. We use this term here, and also the term *communication media*, to refer both to documents and to materials whose messages are primarily visual or auditory.

If you plan to include communication media as a data source in your research, you will need to be systematic about identifying and analyzing them. In the following sections we describe appropriate procedures, depending on whether your research will be conducted from a quantitative or qualitative perspective.

Content Analysis in Quantitative Research

In quantitative research, the analysis of documents typically involves content analysis. Content analysis has been defined as "a research technique for the objective, systematic, and quantitative description of the manifest content of communication."[44] The raw material for content analysis can be any type of document or other communication medium.

Most content analyses in education involve collecting data on various aspects of the messages encoded in the communication product. These analyses generally involve fairly simple classifications or tabulations of specific information. Content analyses of student compositions for language arts classes, for example, could be used to develop a typology of grammatical and spelling errors as well as information on the frequency of different types of errors. This information can be used to revise language arts courses or develop remedial programs.

44. Berelson, B. (1952). *Content analysis in communication research.* Glencoe, IL: Free Press. Quote appears on p. 18.

Steps in Doing a Content Analysis

In describing the steps involved in a content analysis, we refer to a research study of vocabulary instruction in social studies textbooks conducted by Janis Harmon, Wanda Hedrick, and Elizabeth Fox.[45] Although this example involves printed text, the procedures are equally appropriate for the analysis of other communication media, such as Web pages, film, and audio recordings.

Specify research questions, hypotheses, or objectives. The researchers' study grew out of their appreciation of the important role that vocabulary instruction plays in students' ability to understand social-studies concepts presented in the upper-elementary and middle grades. Also, they were aware of effective research-based practices for promoting vocabulary development and wondered whether these practices were represented in the teachers' editions of social-studies textbooks.

These interests and concerns led Harmon, Hedrick, and Fox to frame three research questions to guide their study:

1. What is the nature of the words or key terms selected by the social studies textbook publishers?
2. To what extent and how is vocabulary represented at each grade level and across series of published social studies programs for grades 4–8?
3. What vocabulary instructional supports do publishers provide for teachers?[46]

Textbooks, especially teachers' editions, contain a great many content-related elements. By framing the research questions stated above, these researchers made explicit the scope of their investigation and also focused it sufficiently so that a manageable study could be designed.

Select a sample of documents to analyze. The researchers selected teachers' editions of the social studies textbooks on the 1997–1998 Texas state adoption list for grades 4–8. The teachers' editions were selected because they contain both the text that students read and recommendations to teachers about how to teach this text. It was necessary to have documents that included these recommendations in order to address the third research question stated above.

Develop a category-coding procedure. The essence of a content analysis is the coding of the document's messages into categories. Each category should represent a discrete variable that is relevant to your research objective. The categories should be mutually exclusive, such that any bit of communication can be coded by only category in the category system.

If appropriate, consider employing a coding system that has been used in previous research. This option saves the time required to develop your own system. Also, the use of standard coding categories permits comparison with other studies that have used the same system. Consequently, the research project is more likely to make a contribution to theory and knowledge in the field under consideration.

In the vocabulary study that we are considering here, the researchers developed their own categories, but they represent standard features of textbooks, including teachers' editions. Among the categories are these: instructional objectives; instructional components (i.e., suggestions for tasks that teachers or students might complete in order to enhance instruction and learning); and key terms (i.e., words or phrases highlighted by the

45. Harmon, J. M., Hedrick, W. B., & Fox, E. A. (2000). A content analysis of vocabulary instruction in social studies textbooks for grades 4–8. *Elementary School Journal, 100,* 253–271.
46. Ibid., p. 254.

publishers), with subcategories for (1) general terms that might be found in any subject (e.g., *malady, revenge*), (2) technical multiple-meaning terms (e.g., *table, legend*), (3) technical domain-specific words unique to social studies (e.g., *pueblo, colonist*), and (4) terms relating to specific persons, places, and events (e.g., *Bull Run, Dwight Eisenhower*).

After initial development of the content classification system, you should determine whether several raters can use it with a high degree of consistency. You can do this by calculating a correlation coefficient of inter-rater reliability for different raters' classifications.[47] If the inter-rater reliability is low, you will need to identify points of ambiguity in the content classification system and clarify them. To increase the reliability with which the classification system can be used, it is helpful to develop an explicit set of scoring rules.

An explicit reliability check was not reported by the researchers who conducted the vocabulary study. There are repeated references to "we" in the report, so presumably the researchers checked each others' coding as the content analysis proceeded. Nonetheless, our confidence in the study findings would be strengthened by an explicit reliability check.

Conduct the content analysis. A typical content analysis consists of making a frequency count of the occurrence of each coding category in each document in the sample. The general procedure is to create a computer file that reproduces the text of the document. (You can type the text or possibly use a scanning device that reads the text directly into a computer file.) Then read the computer file and type a code for each message that fits a particular category of the content-analysis system. A computer program can be used to count the frequency of each code or to list together all the text messages that fit a particular code. Computer programs of this type are described in Chapter 14.

The frequency counts can be presented in the results section of the research report. Descriptive statistics also can be reported, for example, the mean number and standard deviation of the occurrences of each coding category across all the documents in the sample. Relationships between the variables represented by the different categories can be analyzed and reported, too.

In the vocabulary study, the researchers conducted several of these statistical analyses. For example, they calculated the mean percentage of key terms that fit each of their categories across the textbooks included in the content analysis. They found that 78 percent of the key terms were in the category of domain-specific words, with another 13 percent in the category of terms relating to specific persons, places, and events. In a related analysis, they asked three teachers to identify what they considered to be key terms in representative samples of text from the textbooks (the samples had been retyped so that the teachers could not identify which terms the publishers had marked as key). Surprisingly, there was only 48 percent agreement between the publishers' and teachers' lists.

Table 9.2 shows the results of an analysis of the instructional components involving key terms that were included in the teachers' editions of the textbooks. As you can see, the researchers classified the instructional components into three categories of tasks: instruction (i.e., teaching activities for the teacher); application (i.e., activities requiring students to use the key term); and review (i.e., activities in the review section of the textbook units, chapters, and lessons). For most of the textbooks, the vast majority of the instructional components involved application and review. A more fine-grained analysis of the components revealed an emphasis on surface-level vocabulary activities, for example, filling in blanks, crossword puzzles, and matching words with their definitions.

47. For information about inter-rater reliability coefficients, see: Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation, 7*(17). Retrieved November 3, 2001 from www.ericae.net/pare/getvn.asp?v=7n=17; Scott, W. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly, 19*, 321–325.

Categories of Vocabulary Tasks in the Teacher's Edition of Social Studies Textbooks

Series	Total	Instruction (%)	Application (%)	Review (%)
Harcourt Brace, grade 4	61	21	31	48
Harcourt Brace, grade 5	8	13	50	38
Harcourt Brace, grade 6	31	26	32	42
Houghton Mifflin, grade 4	64	25	36	39
Houghton Mifflin, grade 5	•46	24	30	46
Houghton Mifflin, grade 6	135	23	36	41
Macmillan, grade 4	60	23	33	43
Macmillan, grade 5	108	27	33	40
Macmillan, grade 6	189	17	54	29
Silver Burdett Ginn, grade 4	40	15	55	30
Silver Burdett Ginn, grade 5	14	0	57	43
Silver Burdett Ginn, grade 6	41	2	49	49
Jarrett, grade 4	7	43	43	14
Benson, grade 7	7	29	0	71
Glencoe, grade 7	21	24	24	52
Holt, Rinehart & Winston, grade 7	6	17	33	50
Holt, Rinehart & Winston, grade 8	31	19	55	26
Prentice Hall, grade 8	17	6	12	82
Scott Foresman, grade 8	8	13	38	50

Source: Table 5 on p. 266 in: Harmon, J. M., Hedrick, W. B., & Fox E. A. (2000). A content analysis of vocabulary instruction in social studies textbooks for grades 4–8. *Elementary School Journal, 100*, 253–271. Copyright 2000 University of Chicago Press.

Interpret the results. The final stage of a content analysis is to interpret the meaning of the results. The interpretive process will depend on the purpose of the study and its theoretical or conceptual framework.

In the vocabulary study, Harmon, Hedrick, and Fox used research knowledge about effective vocabulary instruction as a basis for evaluating publishers' practices as revealed by their content analyses. This research knowledge demonstrates the value of such instructional activities as word sorts, semantic mapping, and graphic organizers to help students make connections between their prior knowledge and new terminology. Harmon, Hedrick, and Fox found virtually no utilization of this research knowledge in the textbooks they analyzed:

> Our findings indicate that although publishers do give consideration to vocabulary in their programs, many activities are still grounded in vocabulary teaching and learning activities that are not supported by empirical evidence. Thus, publishers need to take a more aggressive stance to integrate current knowledge about vocabulary into their instructional procedures.[48]

The researchers also raise an important question for future research, based on their data analysis of teachers' and publisher' identification of key terms in social studies textbooks: "Disagreements between teachers and publishers over what should be highlighted as conceptually loaded terms raises questions concerning how publishers select key

48. Harmon et al., *Content analysis*, p. 269.

terms."[49] This is a good example of the principle that the value of a study often lies less with its findings and more in the questions it raises.

Analysis of Documents and Records in Qualitative Observation

Qualitative researchers often study written communications found in natural situations. These written communications are of various types. Yvonna Lincoln and Egon Guba define **documents** as written communications that are prepared for personal rather than official reasons.[50] In contrast, **records** are written communications that have an official purpose. For example, personal letters, personal diaries, and drafts of articles are documents, whereas legal contracts, commission reports for general circulation, tax statements, and newspaper articles are records.

Documents and records rely primarily on language to convey meaning. Other communications—for example, mathematics, music, and highway signs—rely on different types of sign systems to convey meaning. **Semiotics** is a field of inquiry that studies the meaning of these signs.[51] For example, Peter Manning and Betsy Cullum-Swan did a semiotic analysis of the various meanings conveyed by the different sign systems at McDonald's restaurants, among them, the menu board, lighting, arrangement of space, utensils and food wrappings, outdoor playgrounds, and the use of the "Mc" prefix to label various food items.[52] (Semiotics is explained further in Chapter 15.)

In traditional quantitative research, the meaning of a text is assumed to be invariant across readers and across time. In other words, the meaning is in the text itself, and the meaning can be represented as discrete content variables and studied by the methods of content analysis. In contrast, qualitative researchers believe that the meaning of a text resides in the minds of its writer and its readers. Thus, the meaning of a particular document or record can change from reader to reader and from one historical period to another. Furthermore, a document or record can have different meanings at different levels of analysis. For example, the content of a textbook can be analyzed to determine what topics it covers. However, it also can be seen as a secondary source that is derived from primary sources. (Primary and secondary sources are explained in Chapter 4.) The relationship between the textbook and its primary sources thus can be the focus of a qualitative researcher's study.

To fully understand a document or record, therefore, the qualitative researcher needs to study the context in which it was produced—the author's purpose in writing it, the author's working conditions, the authors intended and actual audience, and the audience's purpose for reading it. The qualitative researcher also must realize that, in reading the text, she creates her own meanings.

Documents and records from the past pose particular problems for the qualitative researcher because it is not possible to interview the author or readers. Also, the researcher cannot observe the situations in which these written communications were used or how they were made available for different audiences. Historians have developed various research methods to deal with these problems. Their methods are described in Chapter 16.

Qualitative researchers follow some of the same steps as quantitative researchers who use text and other communications media as data sources. They typically begin by

49. Ibid., p. 265.
50. Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
51. Manning, P. K., & Cullum-Swan, B. (1994). Narrative, content, and semiotic analysis. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 463–477). Thousand Oaks, CA: Sage.
52. Ibid.

identifying documents and records that are part of the situation that they plan to study. Once they have identified these materials, their next step is to determine which materials might be relevant to their research study. Then they determine how they can collect these materials for analysis within the guidelines for ethical conduct of research. If the materials cannot be removed from the natural situation, it may be possible to make photocopies of them and photographs of them in their setting for later analysis. Otherwise, the researchers will need to devise a method for analyzing them on site. Finally, they will need to consider the validity of the materials. Several conceptions of validity (i.e., internal and external criticism) that apply to documents and records are described in Chapter 16.

The use of documents and records as data sources in qualitative research differs most from quantitative research in the analysis phase. In quantitative research, a set of variables is defined and applied uniformly to all the written communications in the sample. The variables are measured in such a way as to yield quantified data that can be analyzed by conventional statistics. In qualitative research, analysis procedure is likely to be emergent. The same document or record can be analyzed at different points in the study, with each analysis yielding new constructs, hypotheses, and insights. The results of the analysis need not be expressed in quantified form. Furthermore, the same document or record can be analyzed from different perspectives and for different purposes.

The results of the qualitative researcher's analysis take the form of interpretations and hypotheses. Ian Hodder proposed that these hypotheses and interpretations need to be weighed in relation to two different contexts—the context in which the documents and records were developed and the context in which they are now being interpreted for research purposes.[53] The researcher must take into account variations in meaning as they are studied across space, time, and cultures. Hodder suggested five criteria for confirming interpretations based on data obtained from documents and records:

1. internal coherence, meaning that different parts of the theoretical argument do not contradict one other and the conclusions follow from the premises;
2. external coherence, meaning that the interpretation fits theories accepted in and outside the discipline;
3. correspondence between theory and data;
4. the fruitfulness of the theoretical suppositions, that is, how many new directions, lines of inquiry, or perspectives are opened up; and
5. the trustworthiness, professional credentials, and status of the author and supporters of an interpretation.

An example of the use of documents, records, and signs in qualitative research is a study by G. Genevieve Patthey-Chavez concerning her perceptions of the cultural conflict between Latino students and their mainstream teachers in a Los Angeles high school.[54] Patthey-Chavez put the high school in context by describing the neighborhood in which it was located. Her observations included various forms of public communications in the neighborhood:

> Area businesses now sport huge Spanish-language signs next to modest English-language ones; video stores advertise their Spanish-language collections; and several Spanish-language

53. Hodder, Interpretation of documents.
54. Patthey-Chavez, G. G. (1993). High school as an arena for cultural conflict and acculturation for Latino Angelinos. *Anthropology and Education Quarterly, 24*, 33–60.

newspapers and radio and television stations compete very successfully with English-language media.[55]

She concluded from these observations that Salvadorean and Mexican cultural networks have been firmly and successfully established in this neighborhood.

Patthey-Chavez's observations of the school's curriculum focused on the English as a Second Language (ESL) program. In examining the text of ESL curriculum materials in use at the school, she found that they emphasized English grammatical structures. This emphasis, she noted, runs counter to current recommendations for ESL methodology, which emphasizes the development of students' oral communicative proficiency.

Patthey-Chavez drew upon a local newspaper article about the high school to illustrate what she perceived as its assimilationist mission, serving as "a port of entry for educational opportunity and the American way, a port of entry that students, but for a few miscreants, were eagerly passing through."[56] In her research report, Patthey-Chavez illustrated this theme by quoting the following excerpt from the newspaper article:

> Student Celia Toche, 17, came from El Salvador with her family five years ago, knowing no English. Now she is an honor student and hopes to go to college. She is usually so absorbed with physics tests, economic theory, trigonometry exercises and English literature that she rarely thinks anymore of such painful losses and frightening times as when armed guerillas perched on the roof of her school.
>
> Sometimes, however, she is reminded. "My parents didn't get a chance to get an education," she says. "It's a family treasure, to get an education here. I really value it a lot. So when I see people wasting the opportunities they've got here, the choices they have here, that's when I think about El Salvador. Here at Lima [High School], we know what we have—and we appreciate it."[57]

Patthey-Chavez claimed that this student's experience represents a biased view of the high school's effectiveness. In working with other students from El Salvador, she found that some of them perceived El Salvadoran schools to be more rigorous and moral than American schools.

The use of a newspaper article as a data source for this study is significant for two reasons. First, it reveals information about one student's experience at the high school being studied. Second, the newspaper creates meanings that are different from those created if the researcher had reported a personal interview with the student: An interview is a personal event, whereas a newspaper provides a widely circulated public record of a conversation. The fact that the newspaper reporter chose to include this particular conversation supports Patthey-Chavez's claim that the school is supporting an assimilationist mission—a mission that the newspaper reporter apparently supports and wishes the public to know. In fact, one can speculate that the newspaper reporter is telling the readers what they want to hear, namely, that their public schools are striving to assimilate immigrants and succeeding in the task. The inclusion of the newspaper article as a data source widens the context of the study and provides a contrast with Patthey-Chavez's other data that question the assimilationist mission of schools.

55. Ibid., p. 40.
56. Ibid., p. 50.
57. Woo, E. (1986, September 28). Lima High makes a comeback. An inner city shows how it's done. *Los Angeles Times Magazine*, p. 14.
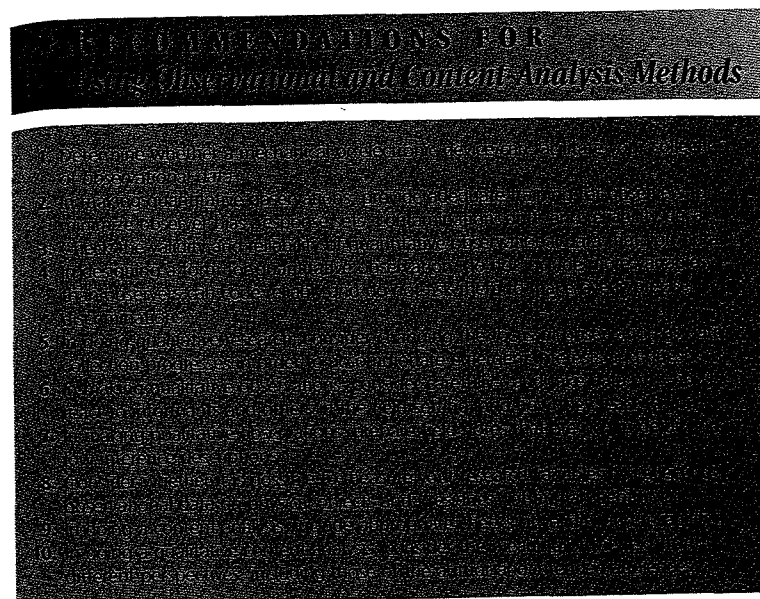
✔ SELF-CHECK TEST

Circle the correct answer to each of the following questions. The answers are provided at the back of the book.

1. Test anxiety is an example of
   a. a descriptive observational variable.
   b. an inferential observational variable.
   c. an evaluative observational variable.
   d. a nonintrusive variable.

2. Recording the number of times that a teacher calls on a student to answer a question about grammar in a Spanish class is an example of
   a. duration recording.
   b. frequency-count recording.
   c. interval recording.
   d. continuous recording.

3. Computers are used in observational research
   a. to time the duration of events.
   b. to detect observer errors in recording observational data.
   c. to make graphical representations of observational data.
   d. all of the above.

4. College students are trained to use a standard schedule to observe children's play behavior. They record their observations on this form, and their data are compared with those of an expert on child psychology. The purpose of this comparison is to check
   a. criterion-related observer reliability.
   b. intra-observer reliability.
   c. inter-observer reliability.
   d. internal reliability.

5. The same observers are asked to record students' grade point average by examining school records and also to observe the students' on-task behavior

in three different classes. This procedure is most susceptible to
a. reliability decay.
b. the error of leniency.
c. observer contamination.
d. observer drift.

6. A researcher observes the meetings of a teachers union. She is a representative of this union, and therefore regularly attends the meetings. The researcher's role is that of
a. complete observer.
b. observer-participant.
c. participant-observer.
d. complete participant.

7. The selected stage of qualitative observation
a. can involve theoretical saturation.
b. must involve a role shift from participant-observer to complete observer.
c. precedes the focused stage.
d. all of the above.

8. Field notes in qualitative research should
a. emphasize high-inference observations.
b. emphasize low-inference observations.

c. avoid personal comments and reflections.
d. have a highly selective focus.

9. Qualitative researchers believe that material objects and residues are
a. expressions of how individuals and groups construct social reality.
b. expressions of the dominant groups within a culture.
c. outside the range of phenomena that they should study.
d. useful to study only if they serve communicative, representational functions.

10. A qualitative researcher is analyzing the textbooks used in present-day Russian schools. To determine the meaning of the textbook content, the researcher is likely
a. to study the authors of the textbooks.
b. to study the teachers who use the textbooks.
c. to study the students for whom the textbooks were written.
d. all of the above.

Quantitative Research Design

You need to design your research study so that it will answer the questions or test the hypotheses you have framed. To accomplish this task, you can refer to standard research designs that quantitative researchers have created and refined over time. These designs are described in Part IV. You need to become familiar with all of them in order to select the most appropriate one for your study and to understand quantitative research studies conducted by others.

Chapter 10 explains descriptive and causal-comparative research designs. A descriptive design is appropriate when your purpose is to create a detailed description of a phenomenon—for example, people's opinions about educational issues or how teachers and students act in certain situations. In a causal-comparative design, the researcher forms two or more groups and compares them in order to explore possible causes or effects of a phenomenon.

Chapter 11 presents the major correlational research designs, which can be used to address questions and hypotheses similar to those addressed in a causal-comparative design. However, correlational research designs involve a different approach to sampling, measurement, and statistical analysis.

The common feature of descriptive, causal-comparative, and correlational research designs is that they involve the study of behavior, cognition, and other attributes of individuals without any intervention by the researcher. In other words, there is no attempt by the researcher to influence the individuals being studied.

In contrast, experimental research designs do involve researcher intervention. The researcher introduces an intervention, such a new educational program or a new teaching method, and observes how the research participants react to it. The major experimental designs are described in Chapters 12 and 13.

As you read the chapters in Part IV, you will find that a major purpose of educational research is to discover and validate cause-and-effect relationships between variables. You also will learn that causal-comparative or correlational research can produce evidence about such relationships, but this evidence generally is not as strong and convincing as that produced by a well-designed experiment.

CHAPTER 14

✦ Case Study Research

OVERVIEW

A good case study brings a phenomenon to life for readers and helps them understand its meaning. We start this chapter by discussing the general characteristics of case studies and the various purposes they serve. The remaining sections of the chapter concern the procedures involved in conducting a case study: formulating a research problem; selecting a case, or cases; defining the researcher's role; collecting and analyzing data; and preparing a report.

OBJECTIVES

After studying this chapter, you should be able to

1. Describe the typical characteristics of a case study.
2. Describe the various purposes that a case study can serve.
3. Explain the meaning of thick description, constructs, themes, patterns, and judgments in case study research.
4. Explain how case study design differs from quantitative research design.
5. Explain how a case study researcher moves from the initial problem statement to finalizing case selection.
6. Discuss the advantages and disadvantages of a multiple-case design as compared to studying a single case.
7. Describe issues that researchers must address in defining their role when conducting a case study.
8. Describe the skills that case study researchers need to function successfully in field sites.
9. State the data-collection methods typically used in case study research.

10. Explain factors to consider in deciding when to stop collecting case study data.
11. Compare the purposes and procedures of interpretational, structural, and reflective data analysis.
12. Compare positivist and interpretivist approaches to determining the validity of case study findings.
13. Describe several methods that can be used to check the validity and reliability of case study findings.
14. Explain how case study researchers deal with the issue of generalizing their findings to other settings.
15. Explain how reflective reporting differs from analytic reporting of case study research.
16. Describe the advantages and disadvantages of case study research.

Introduction

One of the main characteristics of qualitative research is its focus on the intensive study of specific instances, that is *cases*, of a phenomenon. For this reason, qualitative research sometimes is called *case study research*. However, the two terms are not synonymous. Case study research evolved as a distinctive approach to scientific inquiry, partly as a reaction to perceived limitations of quantitative research. Many qualitative researchers subscribe to this approach and use it as a guide to their investigations. Other qualitative researchers subscribe instead to research traditions that generally are compatible with case study methodology, but that have grown up around an interest in particular types of phenomena and the development of special methods for studying them. (We describe these traditions, such as ethnography, semiotics, phenomenology, and historical research, in Chapters 15 and 16.)

Virtually any phenomenon can be studied by means of the case study methods that we describe in this chapter. However, other qualitative research traditions are limited to the study of particular types of phenomena. Thus, in choosing to study a phenomenon, you can choose to do so within the general case study framework described in this chapter, or within a specialized qualitative research tradition. Keep in mind, though, that it is not necessary to adhere to any particular tradition in doing qualitative research. Some qualitative researchers, especially those subscribing to postmodernism, reject any formal methodology and therefore do not align themselves either with case study research or with other qualitative research traditions.

In our view, then, case study research is one of several approaches to qualitative inquiry. We present it as the first of several chapters about qualitative research because of its widespread use in education. Another reason for presenting it first is that elements of the case study approach appear in the specialized qualitative research traditions described in Chapters 15 and 16. Learning the case study approach should make it easier for you to understand the purposes and methods of these other approaches to qualitative research.

Robert Stake observes that, "As a form of research, case study is defined by interest in individual cases, not by the methods of inquiry used."[1] For example, some researchers focus on the study of one case because of its intrinsic interest, whereas other researchers study multiple cases in order to test the generalizability of themes and patterns. The epistemological orientation of most case study researchers is interpretive, but there are some whose methods reflect a positivist orientation. In this chapter, we present the range of methods used in case study research rather than limiting ourselves to the methods of only one epistemological orientation.

Characteristics of Case Studies

To explain the characteristics of a case study, we will use as an example a study reported by Dona Kagan, who collected the data, and four teachers who were themselves the cases studied by Kagan.[2] The study examined the effects of a staff development program on the professional lives of four elementary teachers who participated in it. Kagan conducted and audiotaped a 90-minute interview with each of the four teachers who were in or had recently completed the program.

1. Stake, R. E. (2000). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 435–454). Thousand Oaks, CA: Sage. Quote appears on p. 435.
2. Kagan, D. M., Dennis, M. B., Igou, M., Moore, P., & Sparks, K. (1993). The experience of being a teacher in residence. *American Educational Research Journal, 30,* 426–443.

In the introductory section of their report, Kagan and her colleagues state

> [T]he structure of this article differs from the traditional empirical report by *not* beginning with an elaborate theoretical framework. In the spirit of grounded theory, the teachers' narratives are presented first, then an interpretation is inferred by Dona, a professor of education who worked with the teachers in constructing their stories.[3]

In the following sections, we will analyze this study in relation to four characteristics of case study research: (1) the study of phenomena by focusing on specific instances, that is, cases; (2) an in-depth study of each case; (3) the study of a phenomenon in its natural context; and (4) the study of the emic perspective of case study participants. These characteristics suggest the following definition of **case study research:** it is the in-depth study of instances of a phenomenon in its natural context and from the perspective of the participants involved in the phenomenon.

The Study of Particular Instances

A case study is done to shed light on a **phenomenon,** which is the processes, events, persons, or things of interest to the researcher. Examples of phenomena are programs, curricula, roles, and events. Once the phenomenon of interest is clarified, the researcher can select a case for intensive study. A case is a particular instance of the phenomenon.

In the Kagan study, the phenomenon of interest, in broad terms, was school-university partnerships. In reading the report, we find a more limited phenomenon of interest, namely, school-university partnerships that strengthen the staff development capacity of local school districts. The case selected for study was a particular example of this phenomenon, namely, the Teacher in Residence (TIR) program at the University of Alabama. The program brings experienced elementary teachers to the university for a two-year period to teach and supervise in the university's preservice teacher education program. At the end of their appointment, the teachers return to their districts with "a first-hand knowledge of what would help them design in-service programs."[4]

Any phenomenon has many aspects. Therefore, the researcher will need to select a focus for investigation. The **focus** is the aspect, or aspects, of the case on which data collection and analysis will concentrate. Kagan and her colleagues observed that previous research on school-university partnerships tended to focus on their policies and structures. They decided to choose another focus, namely, "the effects of a school-university staff development program on the professional lives of . . . elementary teachers who participated in it."[5] Program effects, then, comprise the focus of this case study. In other case studies, the focus might be a particular individual, a set of events, a time period, or a component of a program or organization.

In some case studies, it is possible to break down the aspect of the phenomenon on which the case study focuses into *units*. In the Kagan study, four teachers who had participated in the TIR program were selected for detailed investigation. In their report, the effects of the TIR program were presented separately for each of the four teachers, followed by a discussion of themes that were common to all of them. The unit of analysis, then, was the participating teacher, and four such units were studied. Another option would have been to identify school-university partnerships around the United States and select several of them for study. In this option, school-university partnerships would be the unit of analysis.

With the above example in mind, we define a **unit of analysis** as an aspect of the phenomenon that can be sampled, with each member of the sample being studied as a separate case. In the Kagan study, the unit of analysis was teachers in residence. Four such units (i.e., four teachers in residence) were selected for study. Some researchers would call each unit a separate case because each unit is studied intensively. This is acceptable as long as we do not lose sight of the fact that the four cases are part of a larger case that is the main focus of the study, namely, the TIR program at one university.

To illustrate these distinctions further, suppose we are interested in how school leaders facilitate mandated curriculum reforms; therefore, we decide to study a particular curriculum reform mandated by the Oregon Department of Education and implemented by all Oregon school districts. In this example, the *phenomenon* is curriculum reform implementation, the *case* to be studied is the curriculum reform implemented in Oregon, the *focus* is the study of leadership behavior that facilitates curriculum reform implementation, and the *unit of analysis* is Oregon school districts. Given this sampling unit, we might select, let's say, a sample of ten school districts in which to collect data. We could study this phenomenon, though, without defining a unit of analysis. For example, we might decide to study the process by which the Oregon Department of Education developed the curriculum reform. The process might be holistic and embedded in various operations of the Department. Thus, the case would not be reducible to any smaller unit of analysis.

Some researchers select a case for study simply because it is interesting or available. If so, the larger significance of the case might be difficult to discern. However, if the case is conceptualized as an example of a broader phenomenon, the case's significance can be seen in terms of the light it sheds on that phenomenon. In the Kagan study, for example, the findings about the TIR program were considered in terms of their implications for school-university partnerships in general: "Perhaps there is a lesson here for those who are interested in promoting school-university partnerships aimed at enhancing teachers' professional lives . . . "[6] The notion of focus can help you keep in mind that a typical case has many aspects, and that a case study probably will be more manageable and meaningful if you concentrate on just a few of the aspects. Finally, the decision to define a unit of analysis and sample within it can help make your data collection more manageable and yet allow you to make meaningful generalizations from your data analyses.

In-Depth Study of the Case

In a case study, a substantial amount of data is collected about the specific case (or cases) selected to represent the phenomenon. These data are in the form of words, images, or physical objects, although some quantitative data may be collected as well. Often data are collected over an extended time period, and several methods of data collection are used.

To determine the effects of the TIR experience on the lives of the four teachers, Kagan conducted a 90-minute interview with each teacher. The interviews were audiotaped and transcribed. Then the transcriptions were analyzed for the purpose of writing a coherent narrative of the teacher's experience. Afterward, the following occurred:

> Two preliminary drafts of the narrative were returned successively to the interviewee for correction, amendment, and editing: an iterative process of constructing meaning that involved both the interviewer and interviewee in the selection and interpretation of biographical data . . . As coauthors, the four teachers were also invited to modify and edit Dona's commentary, as well as the entire article.[7]

3. Ibid., p. 428.
4. Ibid., p. 427.
5. Ibid., p. 427.

6. Ibid., p. 441.
7. Ibid., p. 428.

From this description we can infer that the narratives were initiated after the initial interviews, but then involved continuing dialogue among the teachers and between the teachers and the principal researcher. The narratives thus reflect in-depth study of the teachers' experience of the TIR program.

Study of a Phenomenon in Its Natural Context

Jerome Kirk and Marc Miller define qualitative research as an approach to social science research that involves "watching people in their own territory and interacting with them in their own language, on their own terms."[8] Typically, but not always, case studies involve fieldwork in which the researcher interacts with study participants in their own natural settings. Even in instances where fieldwork is not done, the goal still is to learn about the phenomenon from the perspective of those in the field.

Robert Yin also emphasizes the importance of studying a phenomenon in its natural context.[9] He observes that case studies typically involve investigation of a phenomenon for which the boundaries between the phenomenon and its context are not clearly evident. Yin argues that these boundaries should be clarified as part of the case study.

In the case study of the TIR program, the teacher narratives ("Mary Beth's Experience," "Karen's Story," "Mary's Reflections," and "Polly's Story") position each teacher's interview statements about their experience in the program within the context of their past and current lives as elementary teachers. For example, the segment on Mary compares Mary's own preservice experience with the current preservice program for teachers, and reveals that her motivation to become a teacher in residence was "to give some realism to the preservice program."[10] The narrative then describes Mary's likes and dislikes about being a teacher in residence and how it influenced her perceptions of classroom teaching and her plans for when she returns to the elementary classroom. Thus, the "boundaries" of this case are Mary's previous and future life as a teacher. Mary's other roles—daughter, parent, spouse, citizen, and so forth—lie outside these boundaries.

Representation of Emic and Etic Perspectives

One purpose of case studies—in some studies, the only purpose—is to develop an understanding of a complex phenomenon as experienced by its participants. This purpose reflects an interpretive (as contrasted with a positivist) stance. As we explain in Chapter 1, interpretive research seeks to study the local, immediate meanings of social actions for the actors involved in them. The researcher's task is to figure out how to view the phenomenon as the participants view it. The participants' viewpoint is called the emic perspective. Typically, the researcher obtains this perspective through direct observation of the participants (sometimes called "insiders") as they behave naturally in the field, and through informal conversations with them. The conversations typically include questions such as, "How did you feel when _____?," "What did you think about when _____?," and "Why did you _____?"

At the same time, case study researchers generally maintain their own perspective as investigators of the phenomenon. Their viewpoint as outsiders, which is called the etic perspective, helps them make conceptual and theoretical sense of the case, and to report the findings so that their contribution to the literature is clear.

While the principal researcher (Kagan) does not refer to her own educational background, a note to the article describes her as a professor of education at George Mason

8. Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative* research. Thousand Oaks, CA: Sage.
9. Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage.
10. Kagan et al., Experience of being a teacher, p. 434.

University and states her specializations as teacher education and school-university partnerships. The research cited in the last section of the report ("Dona's Commentary and Analysis") clearly reflects Kagan's use of her professional experience in the analysis and interpretation of data obtained from the teachers. This is the etic perspective. The close involvement of the four teachers in the data-collection process (each of them was personally involved) and in the reporting of the findings ensures that the emic perspective was represented as well.

Purposes of Case Studies

Researchers generally do case studies for one of three purposes: to produce detailed descriptions of a phenomenon, to develop possible explanations of it, or to evaluate the phenomenon. We describe each of these purposes below.

Description

In a case study whose purpose is description, the researcher attempts to depict a phenomenon and conceptualize it. The depiction can focus on various phenomena, such as: the meanings that the research participants ascribe to their life and environment, contextual factors that influence their life, a series of events and their possible outcomes, and the new or unusual in society. A good depiction will provide what is called a **thick description** of the phenomenon, that is, statements that re-create a situation and as much of its context as possible, accompanied by the meanings and intentions inherent in that situation. The term *thick description* originated in anthropology to refer to a complete, literal description of a cultural phenomenon,[11] but is now used in qualitative research generally. A good example of thick description, in our opinion, is the opening statements in a report about a high school dropout by the educational anthropologist Harry Wolcott:

> "I guess if you're going to be here, I need to know something about you, where you're from, and what kind of trouble you are in," I said to the lad, trying not to reveal my uncertainty, surprise, and dismay at his uninvited presence until I could learn more about his circumstances. It wasn't much of an introduction, but it marked the beginning of a dialogue that lasted almost two years from that moment. Brad (a pseudonym, although as he noted, using his real name wouldn't really matter, since "no one knows who I am anyway") tersely stated his full name, the fact that his parents had "split up" and his mother was remarried and living in southern California, the local address of his father, and that he was not at present in any trouble because he wasn't "that stupid." He also volunteered that he had spent time in the state's correctional facility for boys, but quickly added, "it wasn't really my fault."
>
> It was not our meeting itself that was a surprise; it was that Brad already had been living at this remote corner of my steep and heavily wooded 20-acre home-site on the outskirts of town for almost five weeks.[12]

In creating thick description, the researcher looks for constructs that bring order to the descriptive data and that relate these data to other research findings reported in the literature. A **construct** is a concept that is inferred from observed phenomena and that can be used to explain those phenomena. For example, Wolcott uses the constructs of *education, school, deviant person, deviant act,* and *opportunity* to interpret dropout phenomena as manifested by Brad.

Researchers also can add depth to their descriptions by searching for themes present in the phenomena. We define **themes** as salient, characteristic features of a case. For example,

11. Geertz, C. (1973). *The interpretation of cultures: Selected essays*. New York: Basic Books.
12. Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation.* Thousand Oaks, CA: Sage. Quote appears on p. 68.

if a student—Sally—fails to turn in her homework on time on one occasion, it can be regarded as an isolated event. If Sally habitually fails to turn in her homework when it is due, we can say that this behavior is characteristic of her; it is a noteworthy theme. If Sally also delays doing household chores and getting ready for family outings, we might see a broader theme in her behavior, namely, procrastination.

The Kagan study illustrates the search for themes in case study data, specifically, "themes that could be validated with the professional literature on teaching and teacher education."[13] One of these themes was the functional value of the TIR experience. According to Kagan and her colleagues, the program experience not only gave teachers a different perspective on their own teaching, but served as a form of teacher research, that is, as a way to acquire theoretical rationales for their beliefs and to attach professional labels to them. (Teacher research as a form of action research is explained in Chapter 18.) Another theme was the expansion (not the abandonment) of the role of classroom teacher. The four teachers did not seek the TIR experience for upward mobility from classroom teaching, but rather to gain new insights into their teaching. The researchers described this finding as the most valuable and unexpected insight from their study.

Explanation

Some case study research aims to provide explanations for the phenomena that were studied. We refer to these explanations as patterns, meaning that one type of variation observed in a case study is systematically related to another observed variation. If the researcher does not claim that one variation has a causal effect on the other, we describe it as a relational pattern. If causality is claimed, it is a causal pattern.

An example of a causal pattern can be found in David Thomas's case study of the dynamics of cross-race relationships between protégés and their mentors or sponsors in a large telecommunications company.[14] He interviewed both the senior and junior members of 18 such pairs. In all but two pairs, the senior member was white and the junior member was African-American.

The members of each pair had formed an association in order to develop a relationship that would foster the professional development of the junior member of the pair. Thomas found that the relationships that developed were of two types, consistent with previous research. In a *mentor*-protégé relationship, the senior member provided not only instrumental career support (e.g., advocacy for promotions and performance feedback), but also psychosocial support (e.g., role modeling, counseling, and friendship). In a *sponsor*-protégé relationship, the senior member provided only instrumental career support.

Thomas sought to identify the factors that led to one or the other type of developmental relationship between these cross-race pairs. First, he examined the strategies that individuals used for managing the racial difference between the senior and junior member. Thomas found that individuals used one of two general strategies. The *denial and suppression* strategy involved a preference for not directly addressing the racial difference between the senior and junior member, and the *direct engagement* strategy involved a preference for directly addressing (i.e., communicating about) the racial difference.

Thomas initially expected that the relationship between a junior and senior member would develop into a mentor-protégé relationship only when both parties preferred and used the strategy of direct engagement. Instead, he found that when both parties preferred the same strategy—whether direct engagement or denial and suppression—the relation-

13. Kagan et al., *Experience of being a teacher*, p. 427.
14. Thomas, D. A. (1993). Racial dynamics in cross-race developmental relationships. *Administrative Science Quarterly, 38*, 169–194.

ship between the junior and senior member was more likely to develop into a mentor-protégé relationship than a sponsor-protégé relationship.

Thomas's findings led him to develop a theoretical model of racial dynamics in developmental relationships. The model posits that strategies for managing racial differences have an effect on the type of relationship that develops between the junior and senior member of cross-race pairs. This, then, is a causal pattern: Variations in strategy (denial and suppression vs. direct engagement) are presumed to have an effect on variations in relationships (mentor-protégé vs. sponsor-protégé).

Evaluation

Chapter 17 describes several qualitative approaches to evaluation, including responsive evaluation, fourth-generation evaluation, quasi-legal models of evaluation, and expertise-based evaluation. In each approach, the researcher conducts a case study and makes judgments. In addition, the researcher might create a thick description of the phenomenon being evaluated and identify salient constructs, themes, and patterns. Case studies whose purpose is evaluation are being done with increasing frequency because educational programs that receive government funding are required to undergo formal evaluation.

Designing a Case Study

Standard designs for quantitative research studies have evolved over a period of decades as researchers gained experience in using this approach to inquiry. In planning and executing a new quantitative study, the researcher can look to these standard designs for a suitable "blueprint" of methods and sequential steps that, if employed properly, will yield the desired type of knowledge about the problem being investigated.

This is not true of case study design. Consistent with qualitative research in general, the design of each case study is specific to the phenomenon being studied and to the researcher conducting the study. In Alan Peshkin's view, the essence of case study design is interpretation.[15] It is the researcher's interpretive acts that give "importance, order, and form"[16] to the study. These interpretive acts occur throughout the course of the study:

> Interpretation has to do with the confluence of questions, images, and ideas that are the starting point of my inquiry, or the conceptualizing of my study.
> Interpretation has to do with where I choose to look to see that something is going on with regard to my conceptualization, or the situating of my study.
> Interpretation has to do with the judgment of what to collect that provides documentation for what I think is going on, or the instantiating of my study and the further focusing of its field of inquiry.
> Interpretation has to do with what to select for writing that establishes or affirms what I have identified that has gone on, or the composing of the elements of my research study.
> Finally, interpretation has to do with a perspectival accounting for what I have learned, or the shaping of the meanings and understandings of what has gone on from some point of view . . . [17]

In this view of case study design, the researcher's interpretive skill—acquired through study, apprenticeship, and experience—determines the specific features of the design. In this view, too, case study design is not an event, but a process that occurs throughout the case study.

15. Peshkin, A. (2000). The nature of interpretation in qualitative research. *Educational Researcher, 29*(9), 5–9.
16. Ibid., p. 9.
17. Ibid., p. 9.

Joseph Maxwell presents a more explicit model of case study design, but one which is consistent with Peshkin's emphasis on interpretation.[18] The full design model is shown in Figure 14.1. Its central features are the five components which are shown in circles and linked by lines that illustrate their connectivity: "The components form an integrated and linked by lines that illustrate their connectivity: "The components form an integrated and interacting whole, with each component closely tied to several others, rather than being linked in a linear or cyclic process."[19] The factors surrounding the five components are not linked in a linear or cyclic process."[19] The factors surrounding the five components are not integral to case study design, but rather are influences on, or outcomes of, the design.

The five components in Maxwell's model are not specific methods, but rather sets of issues that are framed as questions. The questions relevant to each component are shown in Table 14.1. If you are planning to do a case study, these questions should prove a helpful aid in creating a design that has "importance, order, and form."

In this and the next sections of the chapter, we present a set of relatively discrete, sequential steps for doing a case study. We take this approach for two purposes: to simplify the process of doing a case study and to provide a basis for comparing case study design with other research designs. In reality, the steps may overlap and be retraced in the course of doing a case study. Also, methods that we describe within steps are not prescriptive. They are subordinate to the larger agenda of respecting the unique features of the case,

Components and Contextual Features of Qualitative Research Design
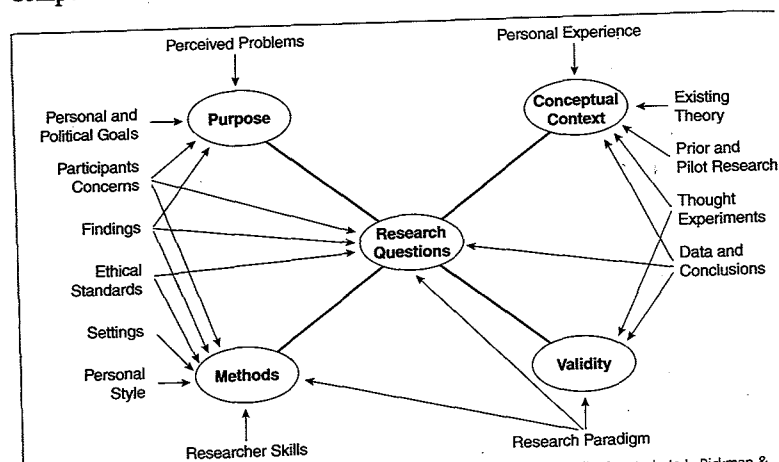
Source: Adapted from Figure 3.2 on p. 73 in: Maxwell, J. A. (1998). Designing a qualitative study. In L. Bickman & D. J. Rog (Eds.), Handbook of applied social research methods (pp. 69–100). Thousand Oaks, CA: Sage. Copyright © 1998 by Sage Publications. Reprinted by permission of Sage Publications, Inc.

18. Maxwell, J. A. (1998). Designing a qualitative study. In L. Bickman & D. J. Rog (Eds.), Handbook of applied social research methods (pp. 69–100). Thousand Oaks, CA: Sage.

19. Ibid., p. 71.

Components of Maxwell's Model of Qualitative Research Design

1. *Purposes:* What are the ultimate goals of this study? What issues is it intended to illuminate, and what practices will it influence? Why do you want to conduct it, and why should we care about the results? Why is the study worth doing?

2. *Conceptual context:* What do you think is going on with the things you plan to study? What theories, findings, and conceptual frameworks relating to these will guide or inform your study, and what literature, preliminary research, and personal experience will you draw on? This component of the design contains the *theory* that you already have or are developing about the setting or issues that you are studying.

3. *Research questions:* What, specifically, do you want to understand by doing this study? What do you *not* know about the things you are studying that you want to learn? What questions will your research attempt to answer, and how are these questions related to one another?

4. *Methods:* What will you actually do in conducting this study? What approaches and techniques will you use to collect and analyze your data, and how do these constitute an integrated strategy?

5. *Validity:* How might you be wrong? What are the plausible alternative explanations and validity threats to the potential conclusions of your study, and how will you deal with these? Why should we believe your results?

Source: Text on p. 71 in: Maxwell, J. A. (1998). Designing a qualitative study. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 69–100). Thousand Oaks, CA: Sage.

keeping interpretation at the center of the case study, and creating a research design that has coherence and value.

Formulating a Research Problem

The first step in planning a case study is to identify a problem that interests you and that is worthy of study. Often the research problem is grounded in the researcher's personal experience with a particular type of student, instructional program, or other phenomenon.

Once identified, the research problem needs to be translated into explicit questions or objectives. The following are two examples from published case studies.

Swidler, S. A. (2000). Notes on a country school tradition: Recitation as an individual strategy. *Journal of Research in Rural Education, 16,* 8–21.

The researcher used an ethnographic, symbolic-interpretive perspective to study a rural one-teacher school in Nebraska.

Research problem: "The present study was undertaken to look at the practices of some of the remaining one-teacher schools, what might be learned from them, and if or how we might capture a glimpse of 'our future in this remaining piece of our past' (Geyer, 1995)." (p. 9).

Goldenberg, C. (1992). The limits of expectations: A case for case knowledge about teacher expectancy-effects. *American Educational Research Journal, 29,* 517–544.

The researcher investigated the academic achievement expectancies of a teacher for two first-grade Hispanic girls in her classroom, the teacher's behavior toward each student, and each student's first-grade reading achievement.

Research problem: "Using a symbolic interactionist perspective (Blumer, 1969, Hewitt, 1984), I try to understand and explain a particular teacher's behaviors toward two children about whom she held diametrically opposed expectations, behaviors that had important consequences for the children's achievement." (p. 522)

Selecting a Case

In this section, we describe in general terms the process of selecting a case or cases. To deepen your understanding of the case selection process, you also should study Chapter 6, which describes many specialized sampling techniques.

The key issue in selecting a case is the decision concerning what you want to be able to say something about at the end of the study. Thus, specifying the case is somewhat comparable to the process in quantitative research of specifying the population to which the results obtained from a sample will be generalized. For example, Sister Paula Kleine-Kracht introduced her study of instructional leadership in a high school with this statement of the research problem:

> There are, however, few examinations of secondary schools in which a principal deliberately chooses to cultivate other members' instructional leadership and decides to personally influence instruction in an indirect way. The following case study is an examination of just such a situation.[20]

In this study, then, the general phenomenon under investigation was instructional leadership. The case was a particular instance (the principal of one high school) of a particular type of instructional leadership, namely, the facilitation of staff members' instructional leadership activities. The findings of the case study might generalize to other principals and might have implications for other manifestations of instructional leadership.

As we explain in Chapter 6, cases in qualitative research are selected by a purposeful sampling process. The particular case to be studied might be selected for various purposes, such as the following: The case is typical; it reflects the phenomenon of interest to an extreme extent; it is a deviant case of special interest; it is politically important.

As we explained, the nature of some cases makes it possible to define a unit of analysis that can be sampled. If this is the situation, the researcher's next decision is to select a sample within the unit of analysis. For example, suppose that the unit of analysis is a group of individuals who share a particular characteristic of interest to the researcher. If the researcher cannot study all these individuals, he will need to consider which of them have experiences or perceptions that give them special value as data sources. There were two such units of analysis in Kleine-Kracht's study of indirect instructional leadership: principals and division chairpersons. Kleine-Kracht selected one school principal from among various principals who potentially could have been studied. The principal's school had five division chairpersons, each responsible for a different part of the school's curriculum. Kleine-Kracht decided to select the entire sample of chairpersons to study, although she could have selected only one or a few chairpersons within this unit of analysis.

Yin proposes that the decision to study multiple cases should be based on replication logic, which is explained in Chapter 6. According to this logic, two or more cases are studied because the researcher predicts the same results for each case (i.e., literal replication), or expects the results to differ for different cases consistent with specific theoretical propositions (i.e., theoretical replication). Our review of published case studies suggests that replication logic is not commonly used, despite its usefulness for testing theoretical propositions.

Although a multiple-case design frequently is used in case study research, Wolcott argues that the study of multiple cases reduces the total attention that can be given to any one of them, and thus serves to weaken rather than to strengthen the study.[21] He expresses

20. Kleine-Kracht, P. (1993). Indirect instructional leadership: An administrator's choice. *Educational Administration Quarterly, 29*, 187–212. Quote appears on p. 187.
21. Wolcott, H. F. (1992). Posturing in qualitative inquiry. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *Handbook of qualitative research in education* (pp. 3–52). San Diego: Academic Press.

a strong preference for studying just one case in depth, especially when the researcher is not experienced in this type of research. Wolcott suggests that more experienced researchers could be left with the responsibility for aggregating and comparing case studies dealing with similar phenomena, and, when appropriate, to discover whatever systematic relationships they may reveal.

The Role of the Case Study Researcher

Quantitative and case study researchers perform similar roles in designing and planning a research study, obtaining necessary institutional review and approval, and obtaining permissions from the site or sites in which the research will be conducted. Once they begin collecting data, however, their roles are quite different. Quantitative researchers specify precise procedures for data collection and analysis. They tend to play a limited role in data collection (e.g., administering questionnaires), or they may even use assistants for this purpose.

The role of case study researchers in data collection is more complex. The researcher is the primary "measuring instrument." This means that she carries out data collection and becomes personally involved in the phenomenon being studied. Thus, the researcher is likely to interact closely with field participants, attend social events in the field setting, and use empathy and other psychological processes to grasp the meaning of the phenomenon as it is experienced by individuals and groups in the setting. Few of these procedures are standardized or can be specified in advance of data collection.

Gaining Entry

Identifying appropriate sites and working with "gatekeepers" to obtain necessary permissions are critical steps in a case study. If not done properly, the researcher may have to abort the study. Also, first impressions created at a site can set the tone for the entire relationship between the researcher and field participants. Issues involved in gaining entry include:

1. identifying people within the field setting with whom to make your initial contact;
2. selecting the best method of communication (e.g., telephone, letter, or personal visit) to deliver your request;
3. deciding how to phrase your request (e.g., focusing on the site's opportunity to contribute to research or on personal benefits to site participants); and
4. being prepared to answer questions and address concerns that might arise both before and after permission is granted.

Margot Ely and her colleagues describe their experiences, and those of their students in a course on qualitative research, in gaining entry and conducting research in field settings.[22] One student's log described her challenges in gaining entry into the field site generally and also into the specific settings she wished to observe and with the specific individuals she wished to interview. The student described her use of the log and her support group (small groups into which students were formed for the duration of the research course) to confront her fears and define her research problem, which involved the experience of aging. She carried out a "dress rehearsal" by visiting a relative who lived in a nursing home, and then identified a different nursing home as a potential site. Unable to find

22. Ely, M., Anzul, M., Friedman, T., Garner, D., & Steinmetz, A. M. (1991). *Doing qualitative research: Circles within circles*. New York: Falmer.

a personal contact who could introduce her, the student made a cold call on the receptionist. She obtained the names of the social service director and the volunteer coordinator, both of whom appeared to be appropriate gatekeepers to approach. Prepared to make a detailed "sales pitch" when she spoke with the director of social work, the student was stunned when she was readily accepted: "You look surprised. Being near a university, we often have students doing research here."[23]

Some of the other students in the qualitative research course had more difficulty gaining entry into the field settings that they initially had selected, and some decided to try elsewhere. Ely and her colleagues note, however: "With a great deal of common sense and sensitivity many researchers can and do turn around rather sticky situations and enter successfully, if more discreetly."[24]

Once you gain entry into a field setting, you will need to enlist the cooperation of the case study participants. Your success will depend on how you present yourself. The student whose nursing home experience we described above recorded this episode in her log:

> If you seem to be fulfilling a role which is aversive to them, friendship and cooperation can evaporate in a flash. I encountered this temporarily when a venerable and very articulate 94-year-old man whom I wanted to interview discovered I had training as a psychologist. He said, "Well, you'll no doubt need to speak to someone else. I'm not a patient." I had to make it clear that I hoped to speak to him as an equal, to be enlightened by his perspective, and not as a psychologist interviewing a patient.[25]

The researcher's appearance also can affect her relationship with the individuals or groups being studied. Consider, for example, the case study by Cathy Evans and Donna Eder involving observations of middle school students during lunch period.[26] (This case study is further described in Chapter 9.) For this phase of their study, the researchers and their assistants sought to fit in with the young people they wanted to observe. Therefore, they dressed and spoke informally and were careful not to react negatively to students' language or behavior, even when they found it disturbing. When interacting with school personnel in other phases of their research, however, the research team dressed and acted as adult professionals.

Ethical Issues

✦ Touchstone in Research

de Laine, M. (2000). *Fieldwork, participation and practice: Ethics and dilemmas in qualitative research.* Thousand Oaks, CA: Sage.

Data collection in case study research poses various ==ethical problems==. An interviewee might experience unexpected emotional difficulty as a result of expressing deeply held and perhaps controversial beliefs and feelings to an interviewer. Individuals might reveal personal information in the presence of a researcher who is in a participant observer role that they would not share with someone perceived to be an outsider. Analysis of personal documents or artifacts might pose ethical issues unless they are willingly surrendered for research purposes. Even when the researcher takes steps to protect the privacy of case study participants, there often are clues in a case study report that make it possible to identify field sites and particular individuals within them.

David Flinders identifies four types of ethics that can provide a basis for viewing and resolving these and other issues that arise in case study research.[27] In ==utilitarian== ethics, re-

23. Ibid., p. 20.
24. Ibid., p. 22.
25. Ibid., p. 25.
26. Evans, C., & Eder, D. (1993). "NO EXIT": Processes of social isolation in the middle school. *Journal of Contemporary Ethnography, 22,* 139–170.
27. Flinders, D. J. (1992). In search of ethical guidance: Constructing a base for dialogue. *International Journal of Qualitative Studies in Education, 5,* 101–115.

searchers judge the morality of their decisions and actions by ==considering the consequences. The most desirable consequence is to produce the greatest good for the greatest number of people==. According to Flinders, utilitarian ethics are difficult to apply in case study research because it is difficult to predict the consequences of a case study while it is in progress.

In ==deontological ethics==, researchers judge the morality of their decisions and actions by referring to ==absolute values==, such as honesty, justice, fairness, and respect for others. Flinders observes that from a utilitarian perspective, deception in case study research could be justified if it could be demonstrated that it did not harm the participants physically or in some other ways. However, deception could not be justified from a deontological perspective. The reason is that deception violates basic values of treating others fairly and with respect.

In ==relational ethics, researchers judge the morality of their decisions and actions by the standard of whether these decisions and actions reflect a caring attitude toward others==. Nel Noddings states that an ethics of caring "takes fidelity to persons as primary and directs us to analyze and evaluate all recommendations in light of our answers to questions concerning the maintenance of community, the growth of individuals, and the enhancement of subjective aspects of our relationship."[28] Relational ethics require, among other things, that the case study researcher be a sensitive, fully engaged member of the participants' community rather than a detached observer.

The fourth basis for considering the morality of a case study is ecological ethics. In ==ecological ethics==, researchers judge the morality of their decisions and actions in terms of the participants' culture and the larger social systems of which they are part. Thus, whereas the other three ethical perspectives consider each case study participant as an individual, ecological ethics consider the participant as a ==member of a larger cultural and social system==. Flinders observes that even a straightforward request such as, "Will you take part in this research?" can mean different things to different people. Depending on the person's cultural background, the request can be perceived as coercive by one individual and as an opportunity to collaborate by another individual. Sex-typed language in a case study report may be viewed as neutral by some readers, but as offensive by others. Data-collection activities may be ethical insofar as the individual participant in a work setting is concerned, but the activities can unfairly disrupt his colleagues' work. An ecological perspective helps avoid these problems by reminding the researcher to consider the larger implications of his local decisions and actions.

Ethical standards for case study research continue to be actively studied and debated. Therefore, if you plan to do case study research, you will need to develop your own ethical perspective. For guidance, we recommend that you review the above ethical perspectives, your institutional review board's standards (see chapter 3), the unique circumstances of the field setting that you wish to study, and your personal values.

Collecting Case Study Data

Case study researchers might begin a case study with ==one method of data collection and gradually shift to, or add, other methods==. Use of multiple methods to collect data about a phenomenon can enhance the validity of case study findings through a process called *==triangulation==*, which we explain later in the chapter.

28. Noddings, N. (1986). Fidelity in teaching, teacher education, and research for teaching. *Harvard Educational Review, 56,* 496–510. Quote appears on p. 510.

The entire range of data collection methods described in Part III can be used in case study research. Also, it is possible to combine qualitative and quantitative methods of data collection. For example, consider Claude Goldenberg's case study of the mismatch between a teacher's expectations and the actual reading achievement of two of her first-grade students.[29] Goldenberg carried out qualitative observation of each child's classroom behaviors, as illustrated by these observational notes in the research report:

> Marta and her group are with the aide, who is giving the children instructions for playing lotto. Marta looks at her lotto cards and smiles. She is sitting quietly. . . . Marta looks at what the boy next to her is doing. . . . Marta then begins to giggle as the aide shows the next picture. She looks at the boy next to her and they giggle together. One girl gives a response to the aide's question and Marta begins to laugh. "Ya, Marta," the aide says. ("Enough, Marta.")[30]

In addition, Goldenberg administered two quantitative measures (standardized tests of reading achievement) to each child.

Data from both types of measures yielded several key findings in this case study. While the teacher initially had low expectations about Marta's academic progress, the teacher told the researcher she spent time with Marta and her mother to encourage greater effort from Marta on reading activity. By year's end, Marta was reading at least at grade level (as measured by the reading tests), and the teacher described Marta as her "wonder child." In contrast, another Hispanic student in the same class, Sylvia, about whom the teacher initially had positive expectations, remained in a low reading group. According to Goldenberg, "in essence, Sylvia stagnated. She was never prompted, as was Marta, to change her work habits."[31]

Personal Involvement in the Data-Collection Process

A type of personal involvement debated among case study researchers is the extent to which they should disclose their personal experiences, feelings, or beliefs to field participants during data collection. For example, when interviewing school personnel about special education policies at the school, is it appropriate for a researcher to discuss her own experience working with special education students, how she felt about this work, and what she believes is the best approach to helping such students? Another issue is the extent to which researchers should include personal experiences, feelings, and beliefs in the research report.

On the issue of personal disclosure to one's case study participants, Daphne Patai describes hearing a well-known historian, Michael Frisch, relate his experience reviewing tapes of research interviews he had conducted.[32] The historian noted that despite all the roadblocks he inadvertently created (which presumably included providing information about himself), most interviewees remained determined to tell him what was important to them and patiently returned to their own themes. This anecdote suggests that ==personal disclosure may have less effect than one might imagine.==

On the issue of personal disclosure in research reports, Patai cites examples of disclosures that she considered excessive or inappropriate:

> Are we really expected to take seriously—and read "generously"—the anthropologist Ruth Behar's claim (in her book *Translated Woman: Crossing the Border With Esperanza's Story*, Bea-

29. Goldenberg, C. (1992). The limits of expectations: A case for case knowledge about teacher expectancy effects. *American Educational Research Journal, 29,* 517–544.
30. Ibid., pp. 528–529.
31. Ibid., p. 537.
32. Patai, D. (1994, February 23). Sick and tired of scholars' noveau solipsism. *The Chronicle of Higher Education,* p. A52.

con Press, 1993) that her struggles to get tenure at an American university should be seen as parallel to the struggles of Esperanza, a Mexican street peddler? Or, to take a different type of example of telling or claiming too much, do readers really benefit from the feminist scholar Nancy K. Miller's description of her father's penis, to which she devotes the closing chapter of her book *Getting Personal* (Routledge, 1991)?[33]

Patai concludes that researchers' personal beliefs and characteristics do not have as much effect on research findings as generally believed, and therefore are better left out of research reports.

In a report discussing the value of sharing one's subjectivity as a researcher, Alan Peshkin offers an opposing view.[34] Peshkin urges researchers to seek out their subjectivity systematically while their research is in progress, so that they can better determine how it might be shaping their inquiry and research outcomes. While carrying out fieldwork in a multi-ethnic high school in a community to which he gave the pseudonym *Riverview,* Peshkin undertook a subjectivity audit of himself. A ==subjectivity audit== involves taking notes about situations connected to one's research that arouse strong positive or negative feelings. The outcome was a list of different aspects of himself, which he described as "discretely characterized I's," reflecting areas in which the researcher's own beliefs and background influenced his perceptions and actions in the research setting. For example, Peshkin discovered the *Ethnic Maintenance I* based on his identity as a Jew, an aspect of his subjectivity that approves of individuals maintaining their ethnicity. Peshkin's *Justice-Seeking I* was aroused by repeated experiences of hearing both residents and nonresidents of Riverview denigrate the community. According to Peshkin, such denigration reflected the fact that until recently Riverview was the only town in a large California county where black people were able to establish residence.

Another issue concerns researchers' personal involvement in their research interpretations, sometimes in the face of objective evidence that their views are incorrect. For example, Deborah Lipstadt, a professor at Emory University, decries what she perceives as a growing assault on historical truth:

> We're in a day and age in which I can make any claim I want. I can say I believe the Buffalo Bills won the Super Bowl. Then I say that it's my opinion and I have a right to it, and you're supposed to back off.[35]

As with most aspects of case study research, there are few firm rules about how much personal involvement or disclosure by a researcher is appropriate. Our view is that if self-disclosure passes a certain point, case study participants and readers of the report will view it as a distraction, or they might question the researcher's qualifications and the trustworthiness of the study's findings. On the other hand, brief comments by the researcher about her background and experiences relevant to the case study may facilitate the reader's understanding of the findings.

Analyzing Data during Data Collection

Data collection is emergent in case study research. ==By this we mean that what the researcher learns from data collected at one point in time often is used to determine subsequent data-collection activities.== Therefore, a case study researcher needs to spend time analyzing the data, at least informally, while data collection still is in progress. Two strategies can facilitate this process: making records of field contacts and thinking "finish-to-start."

33. Ibid.
34. Peshkin, A. (1988). In search of subjectivity—one's own. *Educational Researcher, 17*(7), 17–21.
35. Leo, J. (1994, February 28). The junking of history. *U.S. News & World Report,* p. 17.

Making records of field contacts. Matthew Miles and A. Michael Huberman recommend that case study researchers use standard forms to summarize data-collection events.[36] The completed forms can reveal missing information and thus indicate the need for further data collection. They also can suggest promising directions for subsequent stages of data collection and analysis. One such form is a **contact summary sheet**, on which the researcher summarizes what was learned from each field observation or interview. The form can be predesigned for recording the specific details in which the researcher is interested, for example, the people, events, or situations involved in the contact, the most interesting or problematic aspects of the contact, or ideas about where the researcher should focus attention during the next contact.

Figure 14.2 is an example of a contact summary form used in a case study. The researcher first summarized eight salient points from the contact, which are coded as "Themes/Aspects" in the right column of the figure. Some of these points probably were taken from the researcher's field notes, and others might have come to the researcher's mind while reviewing the notes. In this case, the contact summary sheet was used in a systematic fashion. That is, the researcher treated each point as a data chunk and coded it, using a theme coding system.

The contact summary sheet is not a substitute for the researcher's field notes relating to the field contact. The field notes should be comprehensive and primarily descriptive. In contrast, a contact summary sheet is brief and focuses on what was learned from the field contact that will guide subsequent data-collection activities.

A **document summary form** serves a purpose similar to a contact summary sheet. The researcher writes a brief summary of each document that has been examined, noting the type of document, its uses, a summary of its contents, and ideas about other documents that should be obtained and studied.

Thinking "finish-to-start." Harry Wolcott advises researchers learning how to do qualitative research to work "start-to-finish" but to think "finish-to-start."[37] This approach involves thinking through one's entire research project at the very beginning. Wolcott recommends doing this thinking as soon as a problem has been formulated and necessary agreements have been obtained, but before beginning fieldwork. He suggests that the researcher make tentative decisions at that point about the form in which the completed account will be presented (e.g., monograph, journal article, or project report). Also, the researcher should try to predetermine the relative emphasis that will be given to thick description and to analysis and interpretation of constructs, themes, and patterns:

> if you plan to go heavy on description (a good way to hedge your bet if you entertain doubts about your sophistication at analysis or interpretation), recognize from the outset that rich detail may be critical and you had better not rely solely on "headnotes" for it. Be thinking about possibilities for presenting detailed vignettes and make sure you are recording such events at an adequate level of detail . . . probably including an abundance of direct quotations.[38]

The finish-to-start approach thus helps researchers anticipate the types of data that should be collected, and in what depth.

Ending Data Collection

The decision about when to end the data-collection stage of a case study involves both practical and theoretical considerations. Time and budgetary constraints, or the observa-

36. Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An* expanded sourcebook (2nd ed.). Thousand Oaks, CA: Sage.
37. Wolcott, *Transforming qualitative data*, p. 404.
38. Ibid., p. 404.

Contact Summary Form: Salient Points in Contact, with Theme Codes Assigned

Contact Summary

Type of contact:	Mtg.	Principals	Ken's office	4/2/76	Site	Westgate
		Who, what group	place	date		
	Phone				Coder	MM
		With whom, by whom	place	date		
	Inf. Int.				Date coded	4/18/76
		With whom, by whom	place	date		

1. Pick out the most salient points in the contact. Number in order on this sheet and note page number on which point appears. Number point in text of write-up. Attach theme or aspect to each point. Invent themes where no existing ones apply and asterisk those. Comment may also be included in double parentheses.

Page	Salient Points	Themes/Aspects
1	1. Staff decisions have to be made by April 30.	Staff
1	2. Teachers will have to go out of their present grade-level assignment when they transfer.	Staff/Resource Mgmt.
2	3. Teachers vary in their willingness to integrate special education students into their classrooms—some teachers are "a pain in the elbow."	*Resistance
2	4. Ken points out that tentative teacher assignment lists got leaked from the previous meeting (implicitly deplores this).	Internal Communic.
2	5. Ken says, "Teachers act as if they had the right to decide who should be transferred." (would make outcry)	Power Distrib.
2	6. Tacit/explicit decision: "It's our decision to make." (voiced by Ken, agreed by Ed)	Power Distrib/Conflict Mgmt.
2	7. Principals and Ken, John, and Walter agree that Ms. Epstein is a "bitch."	*Stereotyping
2	8. Ken decides not to tell teachers ahead of time (now) about transfers ("because then we'd have a fait accompli").	Plan for Planning/ Time Mgmt.

Source: Adapted from Figure 4.2 on p. 54 in: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage. Copyright © 1994 by Sage Publications. Reprinted by permission of Sage Publications, Inc.

tion that the participants' patience is running thin, are among the practical considerations that can prompt a decision to end data collection. As to theoretical considerations, Yvonna Lincoln and Egon Guba identify four criteria for determining when it is appropriate to end data collection.[39] The criteria assume that the data have been coded into categories, but

39. Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

they are applicable to other forms of data analysis as well. (Category coding is discussed later in the chapter.) The four criteria are as follows.

1. *Exhaustion of sources.* Data sources (e.g., key informants, institutional files) can be recycled and tapped many times, but at some point it should become clear that little more information of relevance will be gained from further engagement with them.

2. *Saturation of categories.* Eventually, the categories used to code data appear to be definitively established. When continuing data collection produces only tiny increments of new information about the categories in comparison to the effort expended to get them, the researcher can feel confident about ending data collection.

3. *Emergence of regularities.* At some point, the researcher encounters sufficient consistencies in the data that she develops a sense of whether the phenomena represented by each construct occur regularly or only occasionally.

4. *Overextension.* Even if new information still is coming in, the researcher might develop a sense that the new information is far removed from the central core of viable categories that have emerged, and does not contribute usefully to the emergence of additional viable categories.

Analyzing Case Study Data

Data analysis in a quantitative research study is a relatively straightforward process. Suppose the study involves 100 participants and scores on ten variables for each participant—in all, 1,000 bits of numerical data. All these data can be entered into a computer file without much difficulty, and a software program will quickly perform the statistical analyses.

In contrast, even a modest case study will generate a great many pages of observational notes, interview transcripts, and documents obtained from the field setting. Suppose there are 200 such pages, each containing 250 words. That totals 50,000 words. How do you analyze all those words in order to produce significant, meaningful findings?

Renata Tesch reviewed various approaches that have been used to analyze case study data.[40] She classified them into three types: interpretational analysis, structural analysis, and reflective analysis. Each type is explained below. Tesch's typology does not indicate how to analyze case study data at the same level of specificity as a statistics textbook describes how to compute, let's say, a mean score or *t* value. However, it will help you design analysis procedures that are uniquely appropriate to the conditions and purposes of your case study.

Our discussion of data analysis emphasizes verbal data. However, visual images—photographs, drawings, paintings, cartoons, film, and the like—are another important data source in some case studies. While the procedures described below are generally applicable to visual images, you should be aware that qualitative researchers have developed special techniques for their analysis.[41]

Interpretational Analysis

Interpretational analysis is the process of examining case study data closely in order to find constructs, themes, and patterns that can be used to describe and explain the phenomenon being studied. For example, suppose researchers are studying a new U.S. history curriculum. They have available a set of documents written by the curriculum developers (the teacher's edition of the textbook, technical reports, advertisements, etc.), as well as transcripts of interviews with parents whose children are studying the curriculum.

In analyzing these data, suppose the researchers find that both the developers and the parents make frequent reference to the curriculum's supposed goals. Further analysis reveals that two goals are particularly salient to both groups: (1) development of multicultural sensitivity, and (2) development of pride in one's country. One finding of the study, then, is the discovery of these particular goals (which we call *constructs*) as central to this particular curriculum. Suppose further analysis reveals that the curriculum developers most frequently mention multicultural sensitivity as a goal of the curriculum, while parents view national pride as an essential curriculum goal, but one not sufficiently emphasized in this curriculum. This, then, is a discovery about a possibly significant *pattern:* The salience of a particular curriculum goal depends on whether one is a developer or a parent.

Interpretational analysis helps researchers achieve insights such as those in our hypothetical example. The procedures of interpretational analysis can be carried out either manually or by computer. Because of the advantages of computer analysis, we will describe the procedures as they would be carried out with software programs that are available for this purpose.

Segmenting the Database

The first step in interpretational analysis is to compile all the case study data into a computer database. Handwritten notes need to be typed and formatted as computer files. Documents and other previously typed materials can be transformed into computer files by using a computer scanner. Even photographs and other graphic materials can be prepared as computer files in this manner. However, to analyze graphic materials by the procedures described below, you would need to prepare verbal descriptions of their salient features.

The resulting computer database (i.e., all the computer files containing the case study data) now can be manipulated by software programs designed to perform interpretational analyses. The researcher starts by having the software program assign a number to each line of text in the database. Next the researcher breaks the text into meaningful segments. A segment (also called a *meaning unit* or *analysis unit*) is a section of the text that contains one item of information and that is comprehensible even if read outside the context in which it is embedded. For example, in the analysis of interview and questionnaire data, it is common to make each question plus the participant's response a separate segment.

A segment can be any length: a phrase within a sentence, a sentence, a paragraph, or even several pages of text. The researcher identifies each segment by indicating the line number on which it begins and the line number on which it ends.

Developing Categories

One of the most critical steps of interpretational data analysis is developing a set of categories that adequately encompass and summarize the data. The researcher must decide what is worth taking note of in each segment of the database. For example, Michael Ann Rossi did a study of elementary school teachers who had been prepared by staff of the California Mathematics Project to help their colleagues effect reforms in mathematics

40. Tesch, R. (1990). *Qualitative research: Analysis types and software tools.* New York: Falmer.
41. These techniques are described in: Harper, D. (2000). Reimagining visual methods: Galileo to *Neuromancer.* In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 717–732). Thousand Oaks, CA: Sage; Rose, G. (2001). *Visual methodologies: An introduction to interpreting visual objects.* Thousand Oaks, CA: Sage; Van Leeuwen, T. V., & Jewitt, C. (2001). *The handbook of visual analysis.* Thousand Oaks, CA: Sage.

instruction at the local school level.[42] What was of interest to Rossi in her database of interviews, questionnaire responses, and documents was the facilitation strategies that these teacher-leaders used.

To be more precise, we would state that facilitation strategies were a *category* in Rossi's data analysis. A **category** is a construct that refers to a certain type of phenomenon mentioned in the database. (A construct is a concept that is inferred from observed phenomena, for example: self-esteem, cooperation, memory.) A category can be expressed numerically: It is either absent (a value of 0) or present (a value of 1) in any observed phenomenon. This numerical expression is called a *variable* by quantitative researchers, and sometimes by qualitative researchers as well.

Researchers need to develop a category label and definition for each type of phenomenon in the database that is to be analyzed. Also, they need to consider whether a particular category can be analyzed into subtypes. For example, in Rossi's case study, the broad category of *facilitation strategy* was analyzed into 15 subcategories, each of which refers to a different type of facilitation strategy. Subcategories also can represent different degrees or levels of a construct. For example, *perceived helpfulness of a facilitation strategy* might be conceptualized as a category with three subcategories: very helpful, helpful, and not helpful. Subcategories, as we explain below, are useful for detecting relational and causal patterns in case study data.

How do you establish a list of categories for coding the segments in your database? One approach is to use a list of categories developed by other researchers. Rossi used this approach in her study of facilitation strategies. Her list was developed by Matthew Miles and his colleagues for case studies of change agents in education, and subsequently modified by other researchers doing related studies. Examples of the categories are: coaching of individuals, developing a support structure, resource linking, supporting the client emotionally, and training groups.

The other approach is to develop your own categories. You will need to study your data carefully in order to identify significant phenomena, and then determine which phenomena share sufficient similarities that they can be considered instances of the same construct. This construct becomes a category in your category system. You will need to define the category, give it a label, and specify guidelines that you and others can use to determine whether each segment in the database is or is not an instance of the category.

This process of category development is consistent with the principles of grounded theory.[43] Case study researchers who use these principles derive their categories directly from their data rather than from theories developed by other researchers. In other words, the categories are "grounded" in the particular set of data that you collected. Furthermore, the categories seek to explain the phenomena that are observed as well as to describe them. Because of this emphasis on explanation, the categories are considered theoretical. However, even if the categories are purely descriptive, the procedures used in grounded theory are applicable. Therefore, if you intend to develop your own categories, we recommend that you study the principles of grounded theory construction.

Coding Segments

After selecting or developing a category system, the researcher uses it to code each segment in the computer file. It is necessary to examine each segment and decide whether the

42. Rossi, M. A. (1993). The California Mathematics Project: Empowering elementary teachers to be leaders and change agents in mathematics reform. *Dissertation Abstracts International, 54*(09), 3314A. (UMI No. 9305218)
43. Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

phenomenon it describes fits one of the categories in the category system. If it does, the researcher types an abbreviation for the category (e.g., a number or acronym) next to the segment. A segment might contain no instances of any category in one's system, or it might contain instances of several categories, in which case the segment would be coded with the abbreviation for each category.

An example of multiple coding is provided in a case study by one of the authors (J. Gall). She interviewed students enrolled in *Careers Plus*, a program designed to assist people age 50 and over to resume work or change careers. The segment to be coded involved the researcher's question and the student's reply:

> Researcher: At this point, what are your reasons for being in the Careers Plus program?
> Student: To get back to work and to be with people again. I was going to church, but wasn't participating in any activities . . . not a part of. . . . Coming to Careers Plus, I got active. [The instructor] called on me a lot in [class]; I think he liked to see me cry. I'm a crier.

The segment was coded as an instance of four categories:

- Goals/reasons for being in or staying in program: Work
- Goals/reasons for being in or staying in program: Be with people
- Learning activities provided by the program: Career exploration class
- Obstacles to career change: Isolation

This example illustrates that a single segment can provide various types of information of interest to a researcher, and that the information is retrievable through multiple coding.

In the process of coding your segments, you might find that some of your categories are ambiguous or that some segments contain information that is not codable using your category system. If this happens, you will need to revise the category system and then recode all the segments. Researchers typically revise their category system several times before feeling satisfied with it.

Grouping Category Segments

Suppose that the database for a case study contains 500 segments and the category system includes 20 categories (1–20). After coding all the segments, the researcher next would bring together all the segments that were tagged with the Category 1 code. (The process would be repeated for the other 19 categories as well.) Software programs designed for qualitative data analysis can perform this function. For example, if 15 segments were tagged with the Category 1 code, the program would compile all these segments for display on the computer screen, or they can be printed out.

The 15 segments with the Category 1 code are now conveniently grouped together. However, they are decontextualized, that is, they have been removed from their location in the interview transcript, field notes, or other document. This is not a serious problem, however, because each segment includes its line numbers. (You will recall that the software program numbers each line of text in the database.) Thus, you can easily locate each segment in the database of interview transcripts, field notes, and documents. In addition, you might consider developing and applying several categories to help you situate a segment easily after it is removed from its location in the database. For example, suppose the study included five cases (C1, C2, C3, C4, C5), and the data for each case were derived from either an interview with the individual's colleague (COL) or supervisor (SUP), or from observation notes (OBS). Thus, the first segment for Category 1 might have the supplemental context codes (C2, COL), the second segment might be coded (C5, OBS), and so on. In this way, the researcher is easily reminded of the segment's location in the database.

When you examine as a group all 15 segments that were coded with the Category 1 code, you have an opportunity to reconsider whether the construct corresponding to that category is sensible. You might find, for example, that the content of some segments corresponds to the construct as you have defined it, but the content of other segments does not. The solution to this problem is to redefine the construct and perhaps to develop new categories.

In our hypothetical example, the grouping process will yield 20 displays of grouped segments, one for each of the 20 categories. The researcher most likely will print out each display for convenience. Now the displays can be compared to determine whether the categories overlap, whether some categories are confusing or irrelevant to the study, and whether some categories are of particular importance.

Barney Glaser and Anselm Strauss, the developers of the grounded theory approach in qualitative research, coined the term **constant comparison** to refer to this continual process of comparing segments within and across categories.[44] (Keep in mind that each category refers to a separate construct and that, ultimately, what is important is the constructs, not the categories used in data analysis.) The term *constant* highlights the fact that the process of comparison and revision of categories is repeated until satisfactory closure is achieved. Using constant comparison, the researcher clarifies the meaning of each category, creates sharp distinctions between categories, and decides which categories are most important to the study. Although the method of constant comparison refers specifically to the development of constructs that are linked together by a theory, it is applicable to the development of purely descriptive constructs as well.

Strauss and Corbin claim that when using grounded theory principles to determine categories, the researcher should collect data to the point of theoretical saturation. Theoretical saturation occurs when no new data are emerging relevant to an established coding category, no additional categories appear to be necessary to account for the phenomena of interest, and the relationships among categories appear to be well established.

By applying the method of constant comparison, the researcher should arrive at a set of well-defined categories with clear coding instructions. As a final check on the category system, the inter-rater reliability of coding should be determined. (Procedures for computing inter-rater reliability statistics are described in Chapter 9.) Demonstration that the category system can be used with high inter-rater reliability enhances the credibility of the case study findings, and it also encourages other researchers to apply the category system in their own case studies.

Drawing Conclusions

The discovery of constructs in qualitative data can be a significant outcome of a case study. Discovery of themes also is important. (You will recall that earlier in the chapter we defined themes as salient, characteristic features of a case.) For example, Kleine-Kracht used the data from her case study of instructional leadership in a high school (described earlier in the chapter) to identify four themes in effective principals' instructional leadership. One such theme is that "the division chairs are curricular experts."[45] This theme emerged from several sources: data in which the principal stated that he expected the division chairs to be catalysts for instructional leadership and that he had selected them on that basis; a description of the chairs' considerable knowledge of instruction; and data indicating that teachers readily accept the division chairs as instructional leaders.

44. Ibid.
45. Kleine-Kracht, Indirect instructional leadership, p. 209.

If the researcher uses a multiple-case design, the generalizability of constructs and themes across cases can be checked. This process might involve noting whether a particular theme observed in one case also is present in other cases. A more sophisticated check would involve determining whether the particular phenomena that were coded as manifestations of a construct in one case are similar to similarly coded phenomena in other cases.

Multiple-case data also can be analyzed to detect relational or causal patterns. The researcher's constructs can be thought of as variables. Each case can be given a score on each variable, typically: 0 = absent and 1 = present; or 0 = absent, 1 = present to a moderate degree, and 2 = present to a high degree. If the scores on one variable across all the cases systematically covary with scores on another variable, the researcher can infer a relational or causal pattern. Suppose, for example, that the researcher has collected data on ten schools, each constituting a separate case. She finds that the school staffs vary in their confidence in the school administrators, and they also vary in their willingness to try out a state-sponsored educational innovation. Suppose the data analysis reveals that cases (i.e., schools) with a high confidence level tend to have high willingness, whereas cases with a low confidence level tend to have low willingness. Given this result, the researcher would be justified in inferring the following pattern, which possibly is causal: Confidence in school administrators facilitates willingness to experiment.[46]

Patterns can be discovered within a single case as well as in multiple-case analysis. For example, suppose that a researcher observed one child extensively in various settings: a mathematics class, a language arts class, the school playground, at home while doing homework, and at home while playing. The researcher also has collected data on the child's state of mind (thoughts and emotions) in these different settings. These data can be analyzed to determine whether there is a relationship between the settings and the child's state of mind. We might find, for example, that the child feels tense and unfocused in academic work settings, but relaxed and fully engaged while playing. This is a relational pattern within a single case. If the researcher can invoke or develop a theory to explain this relationship, we would characterize it as a causal pattern.

Various software programs for Macintosh and PC computers are available to perform the various data analysis procedures described above. The procedures also can be performed manually. You can write line numbers on the master copy of your interview transcripts, field notes, and other text materials; make a new copy of these materials; mark the new copy into segments; cut out each segment and paste it onto a 3 × 5 card; write a category code (or codes) on each card; group cards having the same category code; and examine the groupings using the method of constant comparison.

These manual procedures will be exceedingly time-consuming if your text materials are extensive. Therefore, many case study researchers plan their data-collection procedures with an eye toward eventually entering them into computer files for manipulation by software designed for interpretational analysis.

Structural Analysis

Structural analysis is the process of examining case study data for the purpose of identifying patterns inherent in discourse, text, events, or other phenomena. To understand how structural analysis differs from interpretational analysis, consider the following example,

46. Other types of procedures that can be used to determine relational and causal patterns in case study data are described in: Miles & Huberman, *Qualitative data analysis.*

which consists of a segment of conversation between a Spanish teacher and one of her students:

Teacher: What does *la casa* mean?
Student: House.
Teacher: That's right. *La casa* means house.

A conversation analyst (see Chapter 15) examining this interaction might note certain features of it, such as:

1. The sequence of speakers within this instructional event was teacher, student, teacher.
2. Each of the teacher's utterances contained more words than the student's utterance.
3. Four Spanish words were uttered.
4. Three words (*la, casa, house*) were uttered twice, and the other six words were uttered once.

The conversation analyst then might examine whether each of these observed phenomena are present in other samples of discourse in this teacher's classroom or in other teachers' classrooms.

This example illustrates the essential feature of structural analysis, namely, that the researcher looks for patterns *inherent* in the data. Very little, if any, inference is required. In contrast, a researcher using interpretational analysis overlays a structure of meaning on the data. For example, suppose the researcher is investigating how students receive feedback in the classroom. The above interaction might be considered a segment, and it might be coded as an instance of feedback because of the utterance, "That's right. *La casa* means house." This classification of the utterance is an inference from the data by the researcher.

Structural analysis is used in conversation analysis, ethnoscience, and other qualitative research traditions. Here are a few examples of the types of educational phenomena that might be investigated in case studies that are based on these traditions: how students' speech patterns change over the course of schooling, the sequence of events in children's stories, how the various parts of textbooks are organized, how curriculum experts conceptualize the high school mathematics curriculum, how teachers and students interact with each other during a lesson, and movement patterns within a school building.

Tesch identified two types of software programs that are useful in doing structural analyses of case study data. Text retrievers are software programs that operate on individual words and fixed sequences of words (e.g., a phrase such as *bill of rights*). These programs can perform such tasks as listing all words in a document, indicating where each word occurs in the document, and counting how many times each word occurs.

Text database managers are software programs that allow the researcher to format a document into fields, code each field, and then retrieve all fields with a given code. For example, if the document is a transcript of a discussion among five educators, each utterance by any of them, no matter how long or short, can be formatted as a field. Next, each field can be assigned one or more codes, such as a code indicating which of the five educators made the utterance. Finally, all fields that have been assigned a given code can be retrieved for display. This feature enables the researcher to retrieve all utterances made by any one of the educators.[47]

47. You may have observed that text database managers are similar to computer programs designed for interpretational analysis. Fields, for example, are similar to segments. However, programs for interpretational analysis have capabilities that text database managers lack.

One particular text database manager described by Tesch has an additional capability. It can search for sequence patterns in a document. Suppose the document is the transcript of the discussion among the five educators mentioned above. This computer program, for example, could identify and compile all sequences in which educator B made a statement that was followed by a statement made by educator D. The identification of such sequences, and their frequency, might reveal interpersonal dynamics that have theoretical or practical implications.

Reflective Analysis

Interpretational analysis and structural analysis involve explicit procedures that are performed in a somewhat prescribed sequence. In contrast, **reflective analysis** is a process in which the researcher relies primarily on intuition and judgment in order to portray or evaluate the phenomena being studied. Terms other than intuition and judgment have been used to describe this process: introspective contemplation, tacit knowledge, imagination, artistic sensitivity, and "examining with a sense of wonder."[48]

Reflective analysis is associated with several qualitative research traditions, including educational connoisseurship and criticism (described in Chapter 17) and phenomenology (described in Chapter 15). We believe, however, that reflective analysis also could be used in case studies that draw on other qualitative research traditions. Its use involves a decision by the researcher to rely on her own intuition and personal judgment to analyze the data rather than on technical procedures involving an explicit category classification system.

Some case studies, especially those associated with ethnographic traditions (see Chapter 14), involve a collaborative effort by a team of researchers. In this situation, the reflective analyses are formed through a dynamic process that is likely to involve conflict, negotiation, and ambiguity and that may result in unusually rich interpretations of the data. Judith Wasser and Liora Bresler suggest the concept of the interpretive zone to characterize this process: "In the interpretive zone, researchers bring together their different kinds of knowledge, experience, and beliefs to forge new meanings through the process of the joint inquiry in which they are engaged."[49]

One way to understand reflective analysis is to compare it with artistic endeavors. The artist reflects on phenomena and then portrays them in such a way as to reveal both their surface features and essences. Many case study researchers engage in similar reflections and portrayals. Reflective analysis is ideally suited for thick description, but it also can lead to the discovery of constructs, themes, and patterns.

The other side of artistic portrayal is criticism. Literary critics, for example, study a literary work in order to develop an appreciation of its aesthetic elements and "message," but also to make critical judgments about its artistic merit. Many case studies conducted by educational evaluators (see Chapter 17) are conducted for similar purposes. These evaluative studies help educators and policy makers understand the features and purposes of educational programs, products, and methods, and also to appreciate their strengths and weaknesses. Just as a literary critic develops reflective ability with experience, so must an educational evaluator build up a store of experience in order to use reflective analysis wisely.

48. All these terms, except the last, are taken from Tesch, *Qualitative research,* p. 69. The phrase, "examining with a sense of wonder," appears in: Barritt, L., Beekman, T., Bleeker, H., & Mulderiz, K. (1985). *Researching educational practice.* Grand Forks, ND: University of North Dakota, Center for Teaching and Learning.
49. Wasser, J. D., & Bresler, L. (1996). Working in the interpretive zone: Conceptualized collaboration in qualitative research teams. *Educational Researcher, 25*(5), 5–15. Quote appears on p. 13.

Because reflective analysis is largely subjective, it is not possible to specify standard procedures for doing this type of data analysis. Apprenticeship with an experienced researcher, followed by considerable practice, is essential. However, a few guidelines from hermeneutical research (see Chapter 15) might be generally applicable. In doing reflective analysis from a hermeneutical perspective, the researcher carefully examines and then re-examines *all* the data that have been collected. As this process continues, certain features of the phenomena are likely to become salient. The researcher then should develop an understanding of these features by themselves and in relation to each other. In other words, the analysis should account for as much as possible of the phenomenon being studied. An interpretation or criticism that fits some of the data should not be contradicted by other data.

Validity and Reliability of Case Study Findings

Case study researchers do not agree in their assumptions about the nature of reality and scientific inquiry. Their different assumptions lead them to hold different views about how to conceptualize and assess the validity and reliability of case study findings.

Positivist Criteria

Some case study researchers subscribe to a positivist philosophy of scientific inquiry, which claims that objective knowledge about the world is possible. In other words, facts about the world are assumed to exist independently of researchers' efforts to know them; if they use the scientific method correctly, they will come to discover those facts. Quantitative researchers also subscribe to these positivist assumptions. Not surprisingly, then, case study researchers with a positivist orientation express a somewhat similar view of validity and reliability as that of quantitative researchers.

Robert Yin exemplifies this type of case study researcher.[50] He judges the quality of case study design by three types of validity criteria and one reliability criterion:

1. **Construct validity** is the extent to which a measure used in a case study correctly operationalizes the concepts being studied.
2. **Internal validity** is the extent to which the researcher has demonstrated a causal relationship between *X* and *Y* by showing that other plausible factors could not have caused *Y*. The criterion of internal validity is not applicable to descriptive case study research because it does not seek to identify causal patterns in phenomena.
3. **External validity** is the extent to which the findings of a case study can be generalized to similar cases.
4. **Reliability** is the extent to which other researchers would arrive at similar results if they studied the same case using exactly the same procedures as the first researcher.

Yin's notions of construct validity and reliability parallel quantitative measurement criteria having the same labels (see Chapter 7). His notions of internal and external validity correspond to the criteria of good experimental design used by quantitative researchers (see Chapter 12).

50. Yin, *Case study research.*

Chain of Evidence

As we explained earlier in the chapter, case study researchers collect raw data in the field, for example, observations of events as they occur. The raw data then are analyzed to yield themes, patterns and causal inferences, which constitute the study's findings. According to Yin, the overall validity of a study is strengthened if the researcher presents a strong chain of evidence, that is, clear, meaningful links between research questions, raw data, and findings. By doing so, the researcher enables the reader to follow the derivation of case study evidence from the initial research questions and its use in the researcher's interpretations.

The researcher should make the chain of evidence explicit in the case study report by providing an audit trail. An audit trail is documentation of the research process followed in the case study. Six types of documentation should be considered for inclusion in an audit trail: (1) source and method of recording raw data, (2) data reduction and analysis products, (3) data reconstruction and synthesis products, (4) process notes, (5) materials relating to intentions and dispositions, and (6) instrument development information. Of course, a case study report would be inordinately long if it included all these materials. However, it may be feasible to include small, representative samples of these materials in a methodology section or appendix. Also, as in quantitative studies, the researcher should hold on to these documentation materials for a period of years after the study so that they can be inspected by other researchers.

Pattern Matching

Some case studies are designed as experiments to test the effects of an intervention, such as a new curriculum, on one or more outcomes, such as student academic achievement and student self-concept. These studies usually are designed within a positivist framework, and thus it is appropriate to speak of independent variables (the new curriculum) and their effect on dependent variables (academic achievement and self-concept).

Yin describes several procedures that can be used to test the internal validity of causal inferences drawn from the findings of this type of case study. The procedures require a set of theoretical propositions that are tested against the case study data. The causal inference is strengthened if **pattern matching** is found, that is, if the patterns discovered in the case study data correspond to predictions drawn from the theoretical propositions. For example, suppose a researcher does a case study of a new curriculum that is based on theories of motivation and self-concept. Because principles derived from the theories are incorporated in the curriculum, the researcher might predict that students who experience the curriculum should derive particular benefits. If the observed pattern of benefits matches those that were predicted, the researcher's inference that the benefits were caused by the curriculum has survived a critical validity check. Yin describes several pattern-matching procedures that can be used, depending on the type of case study data that are available and the type of explanation whose validity is being tested.

Interpretive Criteria

Validity and reliability become problematic if one rejects the positivist assumption of a reality that can be known objectively. How does a researcher arrive at valid, reliable knowledge if each individual being studied constructs his or her own reality (the constructivist assumption), if the researcher becomes a central focus of the inquiry process (the "reflexive" turn in the social sciences), and if no inquiry process or type of knowledge has any authority over any other (the postmodern assumption)?[51]

51. Constructivist and postmodern assumptions and the reflexive turn are explained in Chapter 1.

In considering this question, some researchers have concluded that traditional notions of validity and reliability do not apply to case study data and interpretations. These researchers, whom we refer to as interpretive researchers, instead apply such criteria as plausibility, authenticity, credibility, and relevance. Other interpretive researchers retain the notion of validity, but reconceptualize it using such terms as interpretive validity, catalytic validity, interrogated validity, imperial validity, simulacra/ironic validity, situated validity, and voluptuous validity.[52]

To illustrate the nature of these formulations of validity, we will consider David Altheide and John Johnson's conception of interpretive validity.[53] Interpretive validity refers to judgments about the credibility of an interpretive researcher's knowledge claims. The criteria are of four types, as described below.

1. *Usefulness.* Interpretive research rejects the notion that it is possible to objectively depict the world. Therefore, objectivity cannot be a criterion for judging the validity of interpretive case study findings. In its place, one imposes the criterion of usefulness. One way in which a case study can be useful is that it enlightens the individuals who read the report of its findings. Another way in which it can be useful is that it liberates the individuals being studied, readers of the report, or some other group. Cultural studies, a qualitative research tradition described in Chapter 15, emphasizes this view of usefulness.

2. *Contextual completeness.* In order for case study phenomena to be properly understood, they need to be set within a context. The more comprehensive the researcher's contextualization, the more credible are her interpretations of the phenomena. Altheide and Johnson recommend that case study researchers consider, at a minimum, the following contextual features in interpreting the meaning of the phenomena they investigate: history, physical setting, and environment; number of participants; activities; schedules and temporal order of events; division of labor; routines and variations from them; significant events and their origins and consequences; members' perceptions and meanings; social rules and basic patterns of order.

Altheide and Johnson also emphasize the need for sensitivity to a setting's multivocality and the participants' tacit knowledge in assessing the validity of case study interpretations. Multivocality refers to the fact that, in many settings, participants do not speak with a unified voice. Rather, they have diverse points of view and interests. Case study interpretations are more credible if the researcher demonstrates openness to the possibility of multivocality.

Tacit knowledge refers to the "largely unarticulated, contextual understanding that is often manifested in nods, silences, humor, and naughty nuances."[54] In other words, case study findings are more credible if they incorporate the implicit meanings present in a situation. Implicit meanings are those that the individuals being studied either cannot find the words to express, or that they take so much for granted that they do not explicate them in everyday discourse or in interviews with the researcher.

3. *Researcher positioning.* A researcher's interpretations are more credible and useful if he demonstrates sensitivity in how he relates to the situation being studied. For example, many settings are socially stratified, such as by social class (e.g., upper class vs. lower class) or by role within a work group (e.g., manager vs. production worker). The researcher

52. These types of validity are mentioned in: Altheide, D. L., & Johnson, J. M. (1994). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 485–499). Thousand Oaks, CA: Sage.
53. Ibid.
54. Ibid., p. 492.

should be sensitive to his own various roles (e.g., member of the middle class, researcher employed by a university, participant-observer in the setting) and how they relate to this stratification structure. If a researcher is aware of these role relationships and thoughtfully considers their effects on the study, his findings will have added credibility.

Thomas's study of developmental relationships in cross-race pairs in a corporation, which we described above, illustrates this self-reflective process. Thomas described two major concerns he had about himself in conducting the study. One concern was the potential impact of his being an African-American male of junior rank. His other concern involved his ability to identify and have effective interview rapport with senior managers, particularly white senior managers. In order to manage these concerns, Thomas enlisted two senior white male colleagues who were familiar with race relations research and clinical methods of supervision to serve as his research supervisors.

Some of our beliefs, values, and other personal characteristics may be subconscious and, therefore, may not be amenable to self-reflection. This problem can be addressed by asking colleagues who know you well to review your research project, including its goals, methods of data collection and analysis, and purported findings. They might be able to identify personal characteristics or conditions that threaten the credibility of your findings. Michael Patton provides an example of the need for this process:

> The fieldwork for evaluation of an African health project was conducted over three weeks during which time the evaluator had severe diarrhea. Did that affect the highly negative tone of the report? The evaluator said it didn't, but I'd want to have the issue out in the open to make my own judgment.[55]

You might be able to refute concerns of this sort, thus strengthening the credibility of the findings. On the other hand, you might find it necessary to reconsider your findings.

4. *Reporting style.* The researcher's choice of reporting style can affect the validity of readers' interpretations of the findings. As a researcher reconstructs the participants' phenomenological reality, he must find a way to express this reconstruction in written or graphic form so that it is perceived as credible and authentic. The goal, then, is to achieve verisimilitude, which Patricia Adler and Peter Adler describe as "a style of writing that draws the reader so closely into subjects' worlds that these can be palpably felt."[56] Some case study researchers consider literary structures (e.g., the telling of tales, one-act plays, poetry) to be particularly good formats for achieving verisimilitude.

The following are additional procedures that case study researchers have developed as checks on the validity and reliability of their case study findings. The appropriateness of a specific procedure for a particular study depends on the researcher's philosophy of scientific inquiry and the phenomena being studied. Because validity and reliability are not clearly distinguished in some forms of case study research, we do not present separate procedures for checking validity and reliability. However, those procedures that clearly relate either to validity or to reliability, and not to both, are so indicated.

5. *Triangulation.* Imagine that you are studying youth gangs, and a gang member tells you that he engages in acts of vandalism because he is bored. Is this a valid statement of the respondent's state of mind? If this person also indicates a habitual state of boredom on a structured questionnaire or personality measure, this evidence would strengthen the credibility of his statement. If we find that other gang members make similar statements,

55. Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Thousand Oaks, CA: Sage. Quote appears on p. 472.
56. Adler, P. A., & Adler, P. (1994). Observational techniques. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 377–392). Thousand Oaks, CA: Sage. Quote appears on p. 381.

this finding would be further evidence of credibility. Still another type of evidence would be finding that the statement is consistent with a well-supported theory of aggression.

What has been done in each of these instances is to validate a case study finding by drawing on corroborative evidence. Case study researchers call this process triangulation: It is the process of using multiple data-collection methods, data sources, analysts, or theories to check the validity of case study findings. Triangulation helps to eliminate biases that might result from relying exclusively on any one data-collection method, source, analyst, or theory.

The key to triangulation is to vary in some way the approach used to generate the finding that you are seeking to corroborate. If you generated a finding by a qualitative method, perhaps you can check it by using a quantitative data-collection method. If the finding came from a statement in a group interview by interviewer X, perhaps its validity can be checked by having interviewer Y conduct individual interviews. If you detect a certain construct, theme, or causal pattern in a set of data, perhaps you can check its validity by asking another researcher to review the data independently and see what themes, patterns, or causal explanations she detects. Depending on what type of finding you are seeking to validate and the kinds of resources available to you, other forms of triangulation might be appropriate as well.

Sandra Mathison observed that triangulation in social science research sometimes does not produce convergence, but instead produces inconsistencies or contradictions among findings about the same phenomenon.[57] For example, in studies of controversial, stressful, or illicit phenomena, self-report data might be inconsistent with, or might even directly contradict, data resulting from more direct methods of data collection, such as observation or document analysis. When this happens, it still may be possible to validate the conflicting data by reconciling them within some explanatory framework.

Ovadia Aviram's case study of an Israeli boarding school for Jewish youth illustrates this process of data reconciliation.[58] Aviram's direct observations of how staff members treated and talked about students produced data that were highly incongruent with what staff members reported in interviews and informal conversations with the researcher. Aviram explained these discrepancies as due to the staff's use of "generating appearances," a process of maintaining a necessary illusion to the outside world that the boarding school provides a benign environment for students and staff.

6. *Member checking.* A major purpose of case study research is to represent the emic perspective, that is, reality as constructed by the individuals who were studied. The validity of a researcher's reconstruction of an individuals' emic perspective can be corroborated by member checking, which is the process of having these individuals review statements made in the researcher's report for accuracy and completeness. This was the primary strategy used in the case study of teachers in residence by Dona Kagan and her colleagues to ensure the accuracy of the findings generated by the principal investigator.

Member checking might reveal factual errors that are easily corrected. In other instances, the researcher might need to collect more data in order to reconcile discrepancies. It is possible, too, that the opportunity to read the report will cause participants to recall new facts or to have new perceptions of their situation. The report would then need to be rewritten accordingly.

7. *Outlier analysis.* Rather than ignoring or explaining exceptions away, Miles and Huberman recommend using extreme cases as a way to test and thereby strengthen the

57. Mathison, S. (1988). Why triangulate? *Educational Researcher, 17*(2), 13–17.
58. Aviram, O. (1993). Appearance and reality in a stressful educational setting: Practices inhibiting school effectiveness in an Israeli boarding school. *International Journal of Qualitative Studies in Education, 6,* 33–48.

basic findings: "You need to find the outliers, and then verify whether what is present in them is absent or different in other, more mainstream examples . . . "[59] As we explain in Chapter 5, an outlier is an individual whose score falls at the extreme or end of the score distribution. In case study research, an outlier is an individual or situation differing greatly from most other individuals or situations. For example, in a case study of school innovation, Miles and Huberman found one site where a new practice was evaluated by many teachers as a miraculous cure for local ills. The researchers, however, found two outliers, that is, individuals at the site who had not adopted the practice or had expressed strong criticism of it. Because these individuals had not mastered the innovation as intended and gave reasons for not adopting it that were opposite to those given by adopters, their comments strengthened the validity of the researchers' interpretation that technical mastery of an innovative practice by users leads to positive results.

Miles and Huberman also recommend seeking out individuals who have the most to gain or lose by affirming or denying something. If such individuals give an unexpected answer (e.g., a person who has much to gain by denying a statement affirms it), the researcher can be more confident that they are answering truthfully. A related tactic is to look for negative evidence, that is, to actively seek disconfirmation of what the researcher thinks is true. For example, asking skeptical colleagues to look at one's raw data and independently come up with their own conclusions is a good way to test the soundness of one's analyses and interpretations.

8. *Long-term observation.* Gathering data over a long period of time and making repeated observations of the phenomenon can increase the reliability of case study findings. For example, students' perceptions of school are known to vary depending upon how much of the school year has passed, what the weather conditions are, whether a school holiday is coming or has just passed, and whether they are experiencing personal problems. If data are collected over an extended time period, the researcher might be able to distinguish situational perceptions of school from more consistent trends.

9. *Representativeness check.* To determine whether a finding is typical of the field site from which it is obtained, the researcher should consider whether there was overreliance on accessible or elite informants in collecting data. The researcher also should try to determine how unusual occurrences, or the fact that the researcher was present on some occasions but not others, might have skewed the findings.

10. *Coding check.* In describing interpretational analysis above, we explained how researchers develop or select a category system to code the segments into which interview transcripts, field notes, documents, and other materials have been divided. The reliability of the coding process can be checked using the methods for determining inter-rater reliability developed by quantitative researchers.

Generalizability of Case Study Findings

A research study's findings are generalizable to the extent that they can be applied to individuals or situations other than those in which the findings were obtained. Generalizability is considered an important, achievable goal in quantitative research. In meta-analysis, for example, research findings of various studies on the same phenomenon that were conducted by different researchers are cumulated to determine a mean effect size. This mean effect size is considered to be widely generalizable. (Meta-analysis is explained in Chapter 4.)

59. Miles & Huberman, *Qualitative data analysis*, p. 269.

The generalizability of case study findings is more problematic. Proponents of the grounded theory approach to case study research question the possibility of generalizing findings from the cases that were studied to other cases. Sally Hutchinson, a proponent of grounded theory, expresses this view:

> Is grounded research replicable [i.e., generalizable]? "Probably not." Grounded theory depends on the interaction between the data and the creative processes of the researcher. It is highly unlikely that two people would come up with the exact same theory.[60]

Hutchinson recommends using case study research to develop grounded theory that, in turn, can be used as the basis for quantitative research studies, which she feels are more suitable for testing the generalizability of research findings.

Other researchers, however, believe that case study findings can be generalized. They recommend designing case studies in ways that will increase the probability that the findings will apply to other cases also representing the phenomenon being studied. One approach is to study a case that is typical of the phenomenon. If an atypical case is selected because it is of particular interest, it might be possible to select a typical case for study as well. If a unit of analysis has been defined, a random sample within this unit of analysis could be sampled. For example, if researchers are studying the effects of an experimental instructional method being used in a particular teacher's classroom (the case), they might select a random sample of the teacher's students for intensive data collection and analysis.

Another approach to the issue of generalizing case study findings is to place the responsibility for generalizing on the "consumers" of the findings rather than on the researchers. For example, Sandra Wilson uses the term **reader/user generalizability** to indicate that it is the responsibility of each reader or user of case study research to determine the applicability of the findings in their own situations.[61] Similarly, Lee Cronbach argues that in social science any generalization should be regarded only as a tentative hypothesis that must be tested against the specific conditions operating in each situation.[62]

Researchers can use several strategies to help readers of a case study report determine the generalizability of findings to their particular situation or to other situations. First, researchers should provide a thick description of the participants and contexts that comprise the case, so that readers who are interested in applying the findings can determine how similar they are to the situation of interest to them. Second, researchers should address the issue of whether the selected case is representative of the general phenomenon being investigated. Finally, if a multiple-case design was used, the researchers should conduct a cross-case analysis to help the reader determine whether there was generalizability at least within the cases that were studied.

Reporting a Case Study

Our review of published case studies indicates that they generally follow one of two reporting styles—what we choose to call *reflective reporting* and *analytic reporting*. Elements of either style can appear in the other, but generally the characteristics of one style are dominant. We describe the two styles below, but first we consider the researcher's decision about which case or cases will be reported.

60. Hutchinson, S. A. (1988). Education and grounded theory. In R. R. Sherman & R. B. Webb (Eds.), *Qualitative research in education: Focus and methods* (pp. 123–140). London: Falmer Press. Quote appears on p. 132.
61. Wilson, S. (1979). Explorations of the usefulness of case study evaluations. *Evaluation Quarterly, 3,* 446–459.
62. Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116–127.

Finalizing Definition of the Case

When the researcher has collected "thick" data about various aspects of the phenomenon being studied, it is possible to define the case in different ways. For example, consider Rossi's study of teachers (she called them teacher-leaders) who had participated in the California Mathematics Project. Instead of focusing on the teacher-leaders, she might have focused on data that she had collected about the project (its history, project directors, summer training programs, etc.) and specific facilitation events for which the teacher-leaders were responsible. However, Rossi decided to focus on one particular aspect, namely, the change-agent role of the teacher-leaders. Also, although she started collecting data on three teacher-leaders, she eventually decided to exclude one of them because of unusual circumstances pertaining to that individual. Thus, decisions about which cases and which aspect of the cases to report were not finalized until well after her dissertation proposal had been approved and data collection was underway.

Deciding which case, or cases, to report can be difficult, because technological tools like computers, videotape equipment, and photocopy machines tempt researchers to generate ever more data. The researcher must sort through the data and report only those cases and aspects of them that have the greatest bearing on the questions that interest her. Wolcott makes the point this way:

> The critical task in qualitative research is not to accumulate all the data you can, but to "can" (i.e., get rid of) most of the data you accumulate. That requires constant winnowing, including decisions about data not worth entering in the first place, regardless of how easy that might be to do. The trick is to discover essences and then to reveal those essences with sufficient context, yet not become mired trying to include everything that might possibly be described.[63]

Because case selection can occur late in the research process, Wolcott argues that case study is not a research design or method, but rather an outcome of qualitative research that the researcher chooses at the stage of preparing the report.[64]

Reflective Reporting

As we explained earlier in the chapter, some case study researchers rely heavily or exclusively on reflective analysis of their data rather than on interpretational or structural analysis. These researchers most likely will prefer to write their dissertations or other reports using a reflective reporting style. The two primary characteristics of **reflective reporting** are the use of literary devices to bring the case alive for the reader and the strong presence of the researcher's voice in the report.

Authors of literary works, of course, tell stories. In reflective reporting, we find that the researcher often weaves case study data into a story. It is this type of researcher that Wolcott apparently has in mind when he argues that the ability to be a storyteller, rather than disdain for number crunching, should be regarded as the distinguishing characteristic of qualitative researchers.[65] Among the ways Wolcott suggests for organizing and presenting a case study as a story are: (1) relating events in chronological order; (2) focusing the story on a critical or key event; (3) recounting the events through the eyes of different participants whose perspectives may differ considerably; and (4) reporting a "day-in-the-life," for example, a reconstruction of the first day of fieldwork, or a typical day in the life of a case study participant.

Figure 14.3 presents the headings for Wolcott's case study of a school dropout, which we mentioned earlier in the chapter. The story of Brad (the "sneaky kid") starts in present

63. Wolcott, H. F. (2001). *Writing up qualitative research* (2nd ed.). Thousand Oaks, CA: Sage. Quote appears on p. 44.
64. Wolcott, *Transforming qualitative data.*
65. Wolcott, *Writing up qualitative research.*

Organization of a Published Case Study Involving Reflective Reporting

Adequate Schools and Inadequate Education

The Life History of a Sneaky Kid

*

The Cultural Context of a Free Spirit
The Life History of a Sneaky Kid
   "In the Chute"
      On the loose
      Getting busted
      Second-rate jobs and second-rate apartments
      A new life
      "Picking up" what was needed
      The bicycle thief
      Being sneaky
      I don't have to steal, but
      Breaking and entering
      Inching closer to the chute
      I'm not going to get caught
      Home is the hunter
      Growing up
      Getting paid for dropping out
      Hiding out from life
   Worldview: "Getting My Life Together"
      A job—that's all that makes you middle class
      Building my own life
      Being by myself
      Friends
      I've been more places and done more things
      Some personal standards
      Moderation: Getting close enough, going "medium" fast
      Putting it all together
   Formal Schooling
Adequate Schools and Inadequate Education: An Interpretation Summary

*The report starts with an introductory section that does not have a heading.

Source: Wolcott, F. (1994). *Transforming qualitative data.* Thousand Oaks, CA: Sage. Copyright © 1994 by Sage Publications. Reprinted by permission of Sage Publications, Inc.

time, and then recounts earlier events in his life as appropriate. Much of the story is organized around broader and narrower themes, one of which is "in the chute," a phrase used to describe individuals whose experiences appear to be leading them toward prison. Wolcott states that he arranged the story sequence this way so that "the reader meets Brad on his own ground, first through a recounting of major events and everyday aspects of his life, next through important dimensions of his worldview."[66] The researcher's voice is heard

66. Wolcott, *Transforming qualitative data*, p. 64.

clearly in the last two sections, where Wolcott presents his interpretation and summary of the case. Wolcott states that the reason for this sequencing is that, "By the time readers arrive at the point where *my* thoughts as to what might be done [about school dropouts], I want to be sure they have a sufficient background to form their own assessment of Brad and his circumstances."[67]

Wolcott's life history of a "sneaky kid" illustrates two principles of reflective reporting. First, the organization of the report highlights what the researcher has learned from the data analysis. Second, the researcher keeps the reader in mind in deciding how to present what he has learned.

The researcher can use various literary devices within the framework of storytelling to bring the case alive for the reader and to convey his point of view as the researcher and narrator of the report. Direct quotes of remarks by the case study participants are particularly effective because they clarify the emic perspective, that is, the meaning of the phenomenon from the point of view of the participants. Much of Wolcott's report consists of direct quotes of statements by Brad, for example:

> I guess being sneaky means I always try to get away with something. There doesn't have to be any big reason. I used to tell the kid I was hanging around with, "I don't steal stuff because I need it. I just like to do it for some excitement."[68]

Some case study researchers have used more dramatic methods for conveying their findings. Laura Richardson cites examples of case studies that have been reported in the form of fiction, poetry, drama, oral readings, comedy and satire, and visual presentations.[69] In one of her own studies, she wrote a poem to convey her understanding of an unmarried mother whom she had studied. The poem uses only the mother's language, as arranged by Richardson. The first lines of the poem are as follows:

> The most important thing
> to say   is that
> I grew up in the South.
> Being Southern shapes
> aspirations   shapes
> what you think you are
> and what you think you're going to be.
>
> *(When I hear myself, my Ladybird*
> *kind of accent on tape, I think, OH Lord.*
> *You're from Tennessee.)*[70]

The use of these dramatic forms to represent case study findings has been called the ==performance turn in qualitative research==. Norman Denzin observes that whereas a standard written text involves one-way transmission of findings from the researcher to the reader, the performance turn intends a dynamic interaction between researcher and audience to co-create meaning:

> Through the act of coparticipation, these works bring audiences back into the text, creating a field of shared emotional experience. The phenomenon being described is created through the

67. Ibid.
68. Ibid., p. 77.
69. Richardson, L. (2000). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 923–948). Thousand Oaks, CA: Sage.
70. Richardson, L. (1992). The consequences of poetic representation: Writing the other, rewriting the self. In C. Ellis & M. G. Flaherty (Eds.), *Investigating subjectivity: Research on lived experience* (pp. 125–140). Quote appears on p. 126.

act of representation. . . . A good performance text must be more than cathartic, it must be political, moving people to action and reflection.[71]

Richardson observes that case study researchers who use these unconventional genres typically have a postmodern sensibility. As we explained in Chapter 1, postmodernism casts doubt on all claims to authoritative methods of inquiry and reporting, including mainstream scientific reports. A postmodernist, therefore, would be inclined to view poetry to be just as legitimate a genre for reporting case study findings as the standard journal article format used by quantitative researchers.

Analytic Reporting

An analytic reporting style is appropriate when the researcher has emphasized interpretational or structural analysis of case study data and has conceptualized the study from a positivist or postpositivist perspective. The major characteristics of **analytic reporting** are an objective writing style (i.e., the researcher's voice is silent or subdued) and a conventional organization of topics to be covered: introduction, review of literature, methodology, results, and discussion. This is essentially the same style and organization used to report quantitative research studies. (See Chapter 2 for guidelines for writing quantitative research reports.)

Suppose a case study researcher has done data analyses to identify constructs, themes, and patterns in the phenomena that have been studied. Suppose further that this researcher used a multiple-case design. Should the researcher report the results for each construct, theme, and pattern across all the cases that have been studied? Or should she report each case by itself and show how the constructs, themes, and patterns are manifested in that particular case? If the former approach is used, the construct, theme, or pattern is highlighted; however, the reader does not get a holistic understanding of each case. Conversely, if the latter approach is used, the reader sees each case as a whole, but it is difficult to make cross-case comparisons with respect to particular constructs, themes, or patterns.

We have found that a combination of the two approaches works well. First, the data analysis results for each case are reported, including sufficient thick description so that the participants, events, and context come alive for the reader. Next, a cross-case analysis is given, which notes consistencies and differences in constructs, themes, and patterns across the cases that have been studied.

As in quantitative research reports, tables and figures are an effective way to present the results of case study analyses. A useful sourcebook of various display formats for this purpose was prepared by Miles and Huberman.[72] They distinguish between two types of display formats: a matrix and a network. A **matrix** is a table that has defined rows and columns. Figure 14.4 is an example of a matrix in their sourcebook, taken from a case study of school improvement and change. It represents in summary form the experiences of three teachers who are attempting to implement a new educational practice.

A **network** is a figure for displaying bits of information, each in a separate "node," and links that show how the bits of information relate to each other. Figure 14.5 is an example of a network in Miles and Huberman's sourcebook, taken from case studies of university students. The left part of the figure shows a sequence of experiences of a student who had temporarily left the university. Each box (a "node") contains a separate experience. The right part of the figure shows a different set of nodes, each containing a force or forces moving the student to the next experience.

71. Denzin, N. K. (2000). The practices and politics of interpretation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 897–922). Thousand Oaks, CA: Sage.
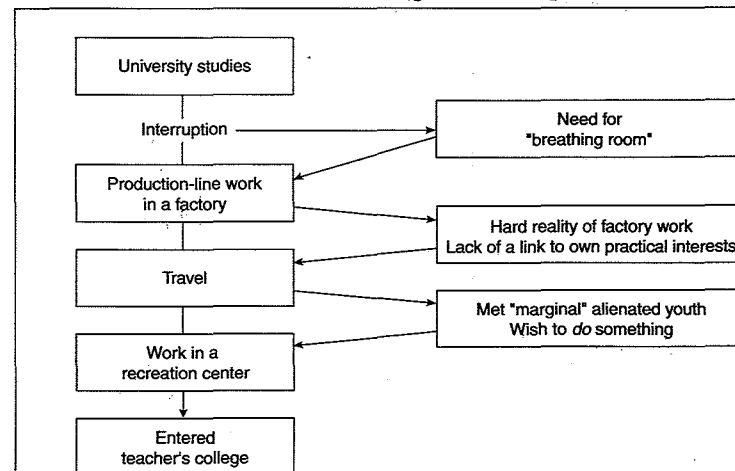72. Miles & Huberman, *Qualitative data analysis*.

Matrix Display of Teachers' Experience in Using a New Educational Practice

User	Feelings/Concerns	How Innovation Looked	What Was User Doing Most?	Problems
Vance	More comfortable with style of teaching and with having kids outside	Still useful, giving good direction and helpful ideas, activities	Working through materials Giving, participating in env'l educ workshops Working with community Off-campus site work	Time too limited for tasks to be done
Drew	Concern with growing number of nonachievers in forestry/ecology class	Too discovery-oriented for kids without biology basics; lecture style more appropriate	Adapting materials and lessons to growing nonachiever population Off-campus site work	Dealing with more nonachievers successfully
Carroll	Excitement with new activities, expanding science program	Same as first year	Working with community Giving, participating in env'l educ workshops Off-campus site work	Overextended activity commitment

Source: Table 7.3 on p. 179 in: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage. Copyright © 1994 by Sage Publications. Reprinted by permission of Sage Publications, Inc.

Event-Flow Network: A Student's Learning and Work Experiences

Source: Adapted from box 5.2 on p. 114 in: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage. Copyright © 1994 by Sage Publications. Adapted by permission of Sage Publications, Inc.

Displays such as those shown in Figures 14.4 and 14.5 can be helpful in two ways. First, they help researchers organize the results of a data analysis and plan the next stage of analysis. Second, displays can be used in the case study report to present research findings so that they are easily comprehended by the reader. Miles and Huberman's sourcebook contains many examples of matrices and networks that are suitable for various purposes in case study research.

## Advantages and Disadvantages of Case Study Research

Now that you have read about case study research, you are in a position to appreciate its advantages over traditional quantitative research methods. One of them is that the case study researcher, through a process of thick description, can bring a case to life in a way that is not possible using the statistical methods of quantitative research. Thus, readers of case study reports may have a better basis for developing theories, designing educational interventions, or taking some other action than they would have from reading only quantitative research reports. Also, thick description helps readers to compare cases with their own situations. These comparisons are more difficult to make when reading reports of quantitative research, which typically provide only statistical analyses and sparse verbal descriptions of the situations that were studied. Furthermore, a good case study report will reveal the researcher's perspective, thus enabling readers to determine whether the researcher has the same perspective on the phenomenon as they do.

Quantitative research designs are well suited for identifying general trends in populations. However, there are situations in which a researcher wishes to learn about a particular individual, for example, an outlier who does not fit the general trend.[73] The case study method is ideally suited to investigating outliers and other unusual phenomena.

Another advantage of case studies is their emergent quality. As researchers collect data and gain insight into particular phenomena, they can change the case on which the study will focus, adopt new data-collection methods, and frame new research questions. In contrast, quantitative research designs are difficult to change once they are set in motion.

The main disadvantage of case studies is the difficulty of generalizing the findings to other situations, although limited generalizations can be made using the procedures that we described in this chapter. Another disadvantage is that ethical problems can arise if it proves difficult in the report to disguise the identity of the organization or individuals that were studied. Also, case studies are highly labor-intensive and require highly developed language skills in order to identify constructs, themes, and patterns in verbal data and to write a report that brings the case alive for the reader. Researchers who lack these resources perhaps can make a better contribution to knowledge in their field of interest by employing quantitative research designs.

73. The outlier phenomenon is discussed in Chapter 15.
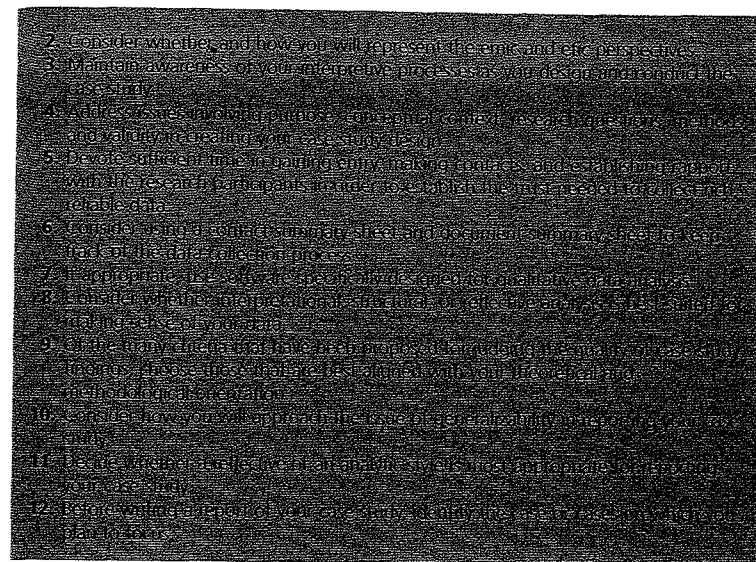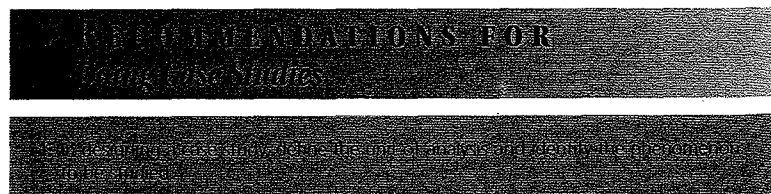
✔ SELF-CHECK TEST

Circle the correct answer to each of the following questions. The answers are provided at the back of the book.

1. The term *emic perspective* refers to
   a. a positivist researcher's view of the phenomenon being studied.
   b. an interpretive researcher's view of the phenomenon being studied.
   c. the research participants' view of the phenomenon being studied.
   d. the reader's view of the phenomenon being studied.

2. Case study researchers need to become personally involved in data collection because
   a. a great deal of data must be collected.
   b. they need to develop a holistic understanding of the phenomenon.
   c. researcher involvement helps ensure the reliability of the findings.
   d. the researcher must continually triangulate the data while it is being collected.

3. The unit of analysis in a case study is
   a. the population from which the case is drawn.
   b. the type of term selected for structural analysis.
   c. the database segment selected for category analysis.
   d. an aspect of a phenomenon that is sampled from possible cases.

4. Experts generally agree that researcher bias in case studies is best handled by
   a. honestly revealing one's possible biases and being willing to have them disconfirmed.
   b. using data-collection methods that rule out the possibility of researcher bias.
   c. using several researchers and seeking consensus in their conclusions.
   d. not studying phenomena in which the researcher has a personal interest.

5. Thick description in case study research refers to
   a. a comprehensive, literal depiction of a phenomenon and its meaning.
   b. description of a phenomenon from both an emic and an etic perspective.
   c. the use of triangulation in writing up case study findings.
   d. reliance on reflective analysis in writing a case study report.

6. In grounded theory, constructs are derived from
   a. a pre-existing theory about the phenomenon being studied.
   b. reflections by the case study participants.
   c. the data that have been collected.
   d. all of the above.

7. The case study finding that there is an association between the amount of structure in teachers' lessons and how teachers think about students' learning processes is an example of a
   a. construct.
   b. theme.

   c. thick description.
   d. pattern.

8. In interpretational data analysis, researcher
   a. searches for the meaning inhe the data.
   b. imposes meaning on the data
   c. searches for naturally occurring ments in the data.
   d. typically uses categories develo other researchers.

9. Usefulness, contextual completene searcher positioning, and reportin are features of
   a. interpretive validity.
   b. case study generalizability.
   c. case study verisimilitude.
   d. audit trails.

10. Reflective reporting of a case stud tends to rely heavily on
   a. presentation of structural data an
   b. an objective writing style.
   c. formats developed for use in rep ing quantitative research studies
   d. the use of literary devices.

Qualitative Research Traditions

OVERVIEW

Qualitative researchers in anthropology, psychology, and other disciplines have developed various ways to study human behavior. Their methods, theories, and accumulated findings constitute distinctive research traditions. In this chapter we explore these traditions and describe how they are used in educational research. Some traditions are particularly well-suited to the investigation of people's inner experience, some to the investigation of social and cultural phenomena, and some to the investigation of communication phenomena such as speech and text. Knowing about these qualitative research traditions will help you think more broadly about a research problem and consider a variety of methods for studying it.

OBJECTIVES

After studying this chapter, you should be able to

1. Describe the characteristics of a qualitative research tradition.
2. Describe the types of phenomena that cognitive psychologists investigate and the research methods that they use.
3. Explain the purpose of phenomenological research and the steps involved in conducting a phenomenological study.
4. Explain the similarities and differences between phenomenography and phenomenological research.
5. Describe the types of educational phenomena that might be investigated by the life history approach.
6. State the goals of the research traditions of symbolic interactionism, event structure analysis, and emancipatory action research.
7. Describe the characteristics of ethnography and the steps involved in conducting an ethnographic study.
8. Explain the strengths and weaknesses of ethnographic research.

9. Explain the goals and underlying assumptions of the cultural studies and critical theory traditions.
10. Describe the types of investigations conducted by researchers who work within the cultural studies and critical theory traditions.
11. Describe the purpose and techniques of ethnomethodological research.
12. State the goals of narrative analysis, ethnoscience, ethnographic content analysis, and the ethnography of communication.
13. Explain the goals, assumptions, and basic concepts of hermeneutical analysis.
14. Explain the goals, assumptions, and basic concepts of semiotic research.
15. Explain the basic principles of structuralism and poststructuralism, and describe the types of phenomena that might be investigated using these approaches.