General Introduction

In this General Introduction we:

- Describe our main goal in the book: helping you select the most effective methods to analyze your data.
- Explain the book's two main organizing questions.
- Discuss what we mean by the remarkably complex term *data*.
- Review the many uses of ordered data, that is, data that have been coded as ranks.
- Discuss the key role of visual/graphic data coding and analyses.
- Consider when the coding process is most likely to occur in your research project.
- Discuss the relation between codes and the world we try to describe using them: between "symbols" and "stuff."
- Present a graphic depiction of the relation of coding to analysis.
- Give examples of the relation of coding to analysis and where to find further discussion of these in the book.
- Look ahead at the overall structure of the book and how you can use it to facilitate your analysis choices.

In this book we give advice about how to select good methods for analyzing your data. Because you are consulting this book you probably already have data to analyze, are planning to collect some soon, or can imagine what you might collect eventually. This means that you also have a pretty good idea of your research question and what design(s) you will use for collecting your data. You have also most likely already identified a sample from which to gather data to answer the research question—and we hope that you have done so ethically.[1] So, this book is somewhat "advanced" in its subject matter, which means that it addresses topics that are fairly far along in the course of a research project. But "advanced" does not necessarily mean highly technical. The methods of

[1]Designs, sampling, and research ethics are discussed in our companion volume, *When to Use What Research Design* (Vogt, Gardner, & Haeffele, 2012).

analysis we describe are often cutting-edge approaches to analysis, but understanding our discussions of those methods does not require advanced math or other highly specialized knowledge. We can discuss specialized topics in fairly nontechnical ways, first, because we have made an effort to do so, and, second, because we emphasize *choosing* various analysis methods; but we do not extensively discuss how to implement the methods of analysis you have chosen.

If you already know what data analysis method you want to use, it is fairly easy to find instructions or software with directions for how to use it. But our topic in this book—deciding when to use which methods of analysis—can be more complicated. There are always options among the analysis methods you might apply to your data. Each option has advantages and disadvantages that make it more or less effective for a particular problem. This book reviews the options for qualitative, quantitative, visual, and combined data analyses, as these can be applied to a wide range of research problems. The decision is important because it influences the quality of your study's results; it can be difficult because it raises several conceptual problems. Because students and colleagues can find the choices of analysis methods to be challenging, we try to help by offering the advice in this book.

If you have already collected your data, you probably also have a tentative plan for analyzing them. Sketching a plan for the analysis before you collect your data is always a good idea. It enables you to focus on the question of what you will do with your data once you have them. It helps ensure that you can use your analyses to address your research questions. But the initial plan for analyzing your data almost always needs revision once you get your hands on the data, because at that point you have a better idea of what your data collection process has given you. The fact that you will probably need to adjust your plan as you go along does not mean that you should skip the early planning phase. An unfortunate example, described in the opening pages of Chapter 1, illustrates how the lack of an initial plan to analyze data can seriously weaken a research project.

WHAT ARE DATA?

What do we mean by **data**? Like many other terms in research methodology, the term *data* is contested. Some researchers reject it as positivist and quantitative. Most researchers appear to use the term without really defining it, probably because a workable definition fully describing the many ways the term *data* is used is highly elusive. To many researchers it seems to mean something like the basic stuff we study.[2] It refers to perceptions or thoughts that we've symbolized in some way—as words, numbers, or images—*and* that we plan to do more with, to analyze further. Reasonable synonyms for *data* and *analysis* are *evidence* and *study*. Whether one says "study the evidence" or "analyze the data" seems mostly a matter of taste. Whatever they are, the data do not speak for themselves. We have to speak for them. The point of this book is to suggest ways of doing so.

[2]Literally, *data* means "things that are given." In research, however, they are not given; they are elicited, collected, found, created, or otherwise generated.

TWO BASIC ORGANIZING QUESTIONS

To organize our suggestions about what methods to use, we address two basic questions:

1. *When you have a particular kind of data interpretation problem, what method(s) of analysis do you use?* For example, after you have recorded and transcribed what your 32 interviewees have told you, how do you turn that textual evidence into answers to your research questions? Or, now that the experiment is over and you have collected your participants' scores on the outcome variables, what are the most effective ways to draw justifiable conclusions?
2. A second, related question is: *When you use a specific method of analysis, what kinds of data interpretation problems can you address?* For example, if you are using multilevel modeling (MLM), what techniques can you use to determine whether there is sufficient variance to analyze in the higher levels? Or, if you are using grounded theory (GT) to analyze in-depth interviews, what kinds of conclusions are warranted by the axial codes that have been derived from the data?

These two questions are related. One is the other stood on its head: What method do you use to analyze a specific kind of data? What kind of data can you analyze when using a specific method? Although the questions are parallel, they differ enough that at various points in the book we stress one over the other. We sometimes address them together, because these two different formats of the question of the relation of evidence and ways of studying it appear often to be engaged in a kind of dialectic. They interact in the minds of researchers thinking about how to address their problems of data interpretation.

Your options for analyzing your data are partly determined by how you have coded your data. Have you coded your data qualitatively, quantitatively, or graphically? In other words, have you used words, numbers, or pictures? Or have you combined these? If you have already coded your data, the ways you did so were undoubtedly influenced by your earlier design choices, which in turn were influenced by your research questions. Your design influences, but it does not *determine*, your coding and analysis options. All major design types—surveys, interviews, experiments, observations, secondary/archival, and combined—have been used to collect and then to code and analyze all major types of data: names, ranks, numbers, and pictures.

RANKS OR ORDERED CODING (WHEN TO USE ORDINAL DATA)

We add **ranks** to the kinds of symbols used in coding because ranks are very common in social research, although they are not discussed by methodologists as much as are other codes, especially quantitative and qualitative codes. Ranking pervades human descriptions, actions, and decision making. For example, a research paper might be judged to be excellent, very good, adequate, and so on. These ranks might then be converted into A, B, C, and so forth, and they, in turn, might be converted into numbers 4, 3, 2, and so forth. If you sprain your ankle, the sprain might be described by a physician as severe, moderate, mild, or with combinations such as "moderately severe." Similar ranks are often used by psychologists describing symptoms. Severity rankings of psychological symptoms or conditions are often based on numerically coded inventories. Ankle sprains are usually judged with visual data; the eye is used to examine an X-ray, a magnetic resonance image (MRI), or even the ankle itself. The arts are no exception to the ubiquity of ranked descriptions; quality rankings by critics of plays, novels, paintings, and so on are routine. In music, composers indicate the tempo at which musicians should play a piece using such ranked tempos as "slowly" (*lento*), "fast—but not too much" (*allegro, ma non troppo*), or "as fast as possible" (*prestissimo*).

Sometimes ranks are given numbers. At other times, numerical continua are divided into categories using cut scores in order to create verbal ranks. Ranks are about halfway between categories and continua. Ranked codes and data can be thought of as a bridge between qualitative categorical codes and quantitative continuous ones. And it is a two-way bridge, with much traffic in both directions. For example, you might describe an interviewee's response to your question by saying that she seemed somewhat hesitant to answer the question—not *very* hesitant or *extremely* hesitant, but *somewhat*. Other interviewees could be described as being willing to answer, whereas still others were eager to do so. If you code your interview responses in this way, you have an implicit or explicit set of ordered categories—or a continuum—in mind. You give those categories (or points on the continuum) labels; they might range from "very eager" to "extremely reluctant" to participate in the interview or to answer particular questions.

Social scientists routinely use concepts and theories based on ranks: psychological conditions, density of social networks, trends in the economy (from mild recession to severe depression), and so on. Ranks are indispensable to social research. Theories,[3] even theories describing relations among quantitatively coded variables, are most often stated in words. Very often the words are descriptions of ranks. Coding using ranks is usually expressed in words or numbers, and it can also be symbolized graphically. Ranked codes are not purely qualitative, quantitative, or visual. Like most codes, they can be arrived at by researchers intuitively and impressionistically or by using fairly strict rules of categorization. Although you have several options when matching concepts to symbols, it is important to be meticulous in recording what you have done in a codebook. It is also important to be certain that you are using analysis techniques appropriate for your codes—for example, different correlations are used for ranked and interval-level data (see Chapter 8).

VISUAL/GRAPHIC DATA, CODING, AND ANALYSES

Visual/graphic data and analyses pervade everything that we write. This is in part because there are so many types and uses of visual/graphic data and analyses. Visual/graphic images can be fairly raw data, such as photographs or video recordings of

[3]We discuss the much-contested term *theory* at several points in the book, most systematically in Chapter 10. Here we can say that a theory is a general description of the relations among variables. An example from social psychology is "expectation states theory": Hierarchies grow up in small groups because of members' expectations of other members' likely contributions to the group's goals.

interviews or interactions. They can be a way to recode other types of data, as when logic models describe a theory of change and a program of action or when bar graphs describe a statistical distribution. And they can be an effective tool of analysis, as when concept maps are used to interpret ideas or when path diagrams are employed to investigate relations among variables. Thus visual/graphic images can be a form of basic data, a way to code data collected in other forms, a way to describe data, and a tool for analyzing them. Although visual/graphic data, codes, and analyses to some extent form a distinct category, they are also discussed in every chapter of this book, because they are indispensable tools for handling and describing one's data as well as for interpreting and presenting one's findings.

A note on terms: We use the terms *visual* and *graphic* more or less interchangeably because that is how they are used in practice by prominent writers in the field. For example, the classic work by Edward Tufte is called *The Visual Display of Quantitative Information*, and his early chapters discuss graphical excellence and integrity. Howard Wainer covers similar topics in *Graphic Discovery*, which recounts several "visual adventures." Nathan Yau's *Visualize This* reviews numerous techniques in statistical graphics, and Manuel Lima's gorgeous *Visual Complexity* mostly uses the term *visual* but calls many of the images he produces *graphs*. Lima pursues the goal of visualizing information—quantitative, qualitative, and visual—which he identifies as the process of "visually translating large volumes of data into digestible insights, creating an explicit bridge between data and knowledge."[4]

AT WHAT POINT DOES CODING OCCUR IN THE COURSE OF YOUR RESEARCH PROJECT?

Although there is no universal sequence, choices about approaches to a research project often occur in a typical order. First, you craft a research question and pick the design you will use to collect the data. The design, in turn, will imply an approach to coding your data. Then your coding choices direct you to some analytical procedures over others. But this order can vary.[5] For example, you may know that your research question requires a particular form of analysis. That form of analysis, in turn, can require that you collect your data and code it in specific ways. For example, if your research question concerns the influence of contexts on individuals' behaviors, you will need to collect data on contexts (such as neighborhoods) and on individuals' behaviors (such as socializing with neighbors, shopping locally, or commuting to work).

Coding data is crucial because an investigation of a research question cannot move ahead without it. When you code your data, you make decisions about how to manage the interface between the reality you are interested in and the symbols you use to think about that reality and to record evidence about it. Two phases are typical in coding.

[4]See, respectively, Tufte (1983), Wainer (2005), Yau (2011), and Lima (2011, quotation on p. 18). A note on footnotes: Based on research with users (graduate students in research methods courses) of books such as this one, we use footnotes rather than in-text citations. For a brief account of that research, see the blog entry "Citation Systems: Which Do You Prefer?" at *http://vogtsresearchmethods.blogspot.com*.

[5]For further discussion, see the Introduction to Part I.

First you define your concepts[6] specifically enough to identify relevant phenomena and collect relevant data. Second, you assign values, such as names or numbers, to your variables in order to prepare them for analysis.[7] The first step in coding is to decide how you will identify your variables (a.k.a. attributes) in order to collect data: Is this a neighborhood? What are its boundaries? The second step is deciding on the coding symbols you will use to produce values you can use in your analyses: Is this neighborhood densely populated? Are particular instances of socializing in the neighborhood organized or spontaneous? The coding symbols can be pictures,[8] words, numbers, ranks, or some combination of these.

CODES AND THE PHENOMENA WE STUDY

Whatever coding scheme you use, a fundamental question is the relation between the symbols and the phenomena they represent. Linguistic philosophers have called the relation between reality and the symbols we use to express it "words and the world."[9] We think of the relationship more broadly to include numbers and pictures as well as words; in our shorthand we call it "symbols and stuff," or, more formally, representations and realities. The key point is that without symbols, you can't study "stuff." The symbols you choose surely influence your understanding of stuff, but not in ways that can be easily specified in advance. The quality of the symbols, their validity, importantly determines the quality of any conclusions you draw from your data.[10]

Most research projects can, and frequently should, involve coding, and therefore analysis, with all three major types of symbols: quantitative, qualitative, and graphic or visual (such as color coding). Often, in any particular project, one of these will be the dominant mode of coding and analysis, but the others generally have a valuable, and perhaps unavoidable, role. Our own beliefs about using multiple forms of coding and analysis are not quite uniform. Our opinions range from the hard position that "it is impossible to think about anything important without using all three" to the softer "there are often many advantages to combining the three in various ways." Although we don't want to digress into epistemology or cognitive psychology, we think that hard and fast distinctions between verbal, numerical, and graphical symbols are difficult to maintain and not particularly useful.[11] In most studies we have conducted, we have

[6]These definitions are often called operational definitions by researchers collecting quantitative data. Fuller discussion of these terms can be found in relevant sections of this volume.

[7]These processes have been described several ways, and different methodologists prefer different terms. For example, some qualitative researchers resist the term *variables* for the things they study; others think that the term *coding* is inappropriate. Helpful descriptions of the processes of coding concepts from different perspectives are given by Jaccard and Jacoby (2010) on the more quantitative side and by Ragin (2008) on the more qualitative.

[8]Network diagrams might be especially useful for this example. For an overview, see Lima (2011) and Christakis and Fowler (2009). Genograms could be even more useful; see Butler (2008).

[9]The classic texts are Austin (1962) and Searle (1969).

[10]For a discussion of valid data coding, see the Introduction to Part I of this book and the Conclusion to Vogt et al. (2012).

[11]See Sandelowski, Voils, and Knafl (2009) on "quantitizing."

combined them. Sometimes we have used formal techniques of mixed method analysis to devise common codes for verbally and numerically coded data. More often we have used graphic, verbal, and numerical data coding sequentially to build an overall interpretation.

Because we think that combined or mixed data are so often helpful for effective analysis and interpretation, we discuss multimethod research throughout this volume rather than segregating it in a separate part of the book.[12] The examples of coding and analysis recounted in the upcoming section drive home the point by illustrating how natural it is to move from one form of coding and analysis to another as you traverse a research project and to unite them in an overall interpretation.

A GRAPHIC DEPICTION OF THE RELATION OF CODING TO ANALYSIS

The typical sequence in a research project leads from coding to analyses. This is illustrated in Figure 1, which also describes how we organized our thinking as we wrote this book. We look at coding and choices among verbal, numerical, graphic, and combined codes (see the left side of the figure; discussed in Part I) and then we review choices among qualitative, quantitative, graphic, and combined modes of analysis (see the right side, as discussed in Parts II and III). Please note that this figure should *not* be read to imply a necessary thematic unity of coding types and analysis methods. It may be more common for attributes coded with words to be analyzed qualitatively or for variables coded with numbers to be analyzed quantitatively, but this is a tendency, not a logical entailment. Researchers have more choices than would be the case were these relations between codes and analyses logical necessities. Because they are not necessary relations, the burden of choice—or, more positively, the freedom to choose—is great.

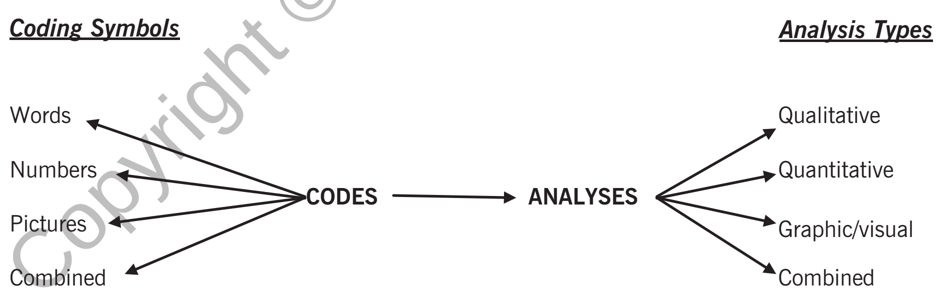

FIGURE 1. The relation of coding to analysis. (*Note.* For an explanation of why the arrows in the figure point in the directions they do, see the discussions of factor analysis [FA] and principal components analysis [PCA] in Chapter 9. The figure is modeled after FA, not PCA.)

[12] The one exception is Chapter 13, in which we address some of the more technical considerations in combing data that have been coded in different ways.

EXAMPLES OF CODING AND ANALYSIS

Rather than continuing to discuss coding and analysis abstractly, we present some brief examples of approaches that one could take to data coding and analysis. There is one set of examples for each of the chapters on coding, and these are tied to relevant chapters on analysis. Each brief example illustrates the interaction between selecting coding and analysis methods and how effective choices can lead to compelling interpretations of your data.

Example 1: Coding and Analyzing Survey Data (Chapters 1 and 8)

Although surveying is usually considered a method of collecting and analyzing quantitative evidence, this is a simplification. Say that you are conducting survey research to investigate attitudes. You collect data about each of the attitudes. But what are attitudes? They are theoretical constructs expressed in *words*. To study them, you could ask respondents to react to statements about attitudes by picking options on a Likert *ranking* scale, which typically uses the following *words*: *strongly agree, agree, neutral, disagree*, and *strongly disagree*. At this point you might assign *numbers* to those words: 5, 4, 3, 2, and 1 are typical. Once numbers are assigned to the words on the scale, you can use quantitative techniques, such as factor analysis, to see whether the items in your presumed scale actually hang together. Using that *quantitative* method, which usually employs *graphic* techniques (such as scree plots), you may find that the items actually form two quite distinct numerical scales. You label those quantitative scales using *words* to identify your new theoretical constructs.[13] This example illustrates how it can be nearly impossible to avoid applying qualitative, quantitative, ranked, and graphic coding and analysis to the same research problem. It also illustrates the pervasiveness of mixed or combined methods of coding and analysis and why we discuss them in *every* chapter of the book.

Example 2: Coding and Analyzing Interview Data (Chapters 2 and 11)

Say you are interviewing people to ask them about their reactions to a social problem. Your main method of data collection is verbal interaction, which you audio- and videotape. You make a transcript of the words, which you analyze using textual techniques. Using your audio or video recording, you analyze gestures, tones of voice, pauses, and facial expressions. You might count and time these (as numbers) or assign ranked verbal codes, such as *strong, moderate*, and *weak* reactions, which you then enter into your notes. You might use grounded theory for the analysis of transcripts, or one of the more quantitative forms of content analysis, or one of the qualitative computer packages (such as Ethnograph) to help you organize and analyze your data.[14] And you might combine these with one of the more quantitative approaches to textual analysis. This example

[13] For an example of this kind of coding and analysis, see Vogt and McKenna (1998).

[14] Some grounded theory researchers embrace computer packages; others reject them; see Chapter 11. The old standbys remain a good place to start when coding interview data (Miles & Huberman, 1994; Spradley, 1979).

illustrates the wide range of choices open to researchers, as well as, again, the pervasiveness of opportunities to apply combined or mixed methods of analysis.

Example 3: Coding and Analyzing Experimental Data (Chapters 3 and 7)

Experiments have a prominent place in most lists of quantitative methods. But the interventions or treatments in experimental social research are not usually quantitative, although they are often coded with a 1 for the experimental group and a 0 for the control group. Here are three quick examples of experimental research and the wide range of coding and analysis methods that can be applied to experimental data. In a survey experiment,[15] respondents were shown two versions of a video depicting scenes of neighbors interacting; the scenes were identical except that the actors in the two videos differed by race. Respondents answered survey questions in which they rated the desirability of the neighborhoods; their ratings were coded with a rank-order variable and analyzed quantitatively. Race importantly influenced individuals' ratings of neighborhood desirability.[16] Another example is a study of the so-called Mozart effect (that listening to Mozart supposedly makes you smarter). The treatment was listening to different types of music (or other auditory phenomena). The dependent measure was obtained with a nonverbal (progressive matrices) IQ test, which resulted in a numerical score. Listening to Mozart had no effect.[17] As a final example, Kahneman discussed studies in which participants briefly looked at photos of political candidates to judge their "trustworthiness." Trustworthiness was coded *verbally* and was associated with other *visually* observed traits (e.g., type of smile). Those qualitative, verbal judgments of visual phenomena were good predictors of election results; that is, they were used in *quantitative* analyses of voting outcomes.[18]

Example 4: Coding and Analyzing Observational Data (Chapters 4, 11, and 12)

In observational studies of organizations, fieldnotes and documents can be used to collect and code data on quality, duration, and number of interactions of members of the organization. Sociograms or other graphic depictions of interactions among people in the organization's networks might be constructed.

For example, in her study of novice teachers, Baker-Doyle investigated each of her participants' social and professional support networks, and she coded these as network diagrams.[19] With these network diagrams, she was then able to characterize the social capital of individual teachers and to come to some useful conclusions about helping new teachers to be successful. The network diagram is becoming a common way to

[15] See Chapter 3 for a discussion of this method.

[16] Krysan, Couper, Farley, and Forman (2009).

[17] Newman et al. (1995).

[18] Kahneman (2011); see especially pages 90–91.

[19] Baker-Doyle (2011).

code interactions of all kinds as a means of understanding human social capital and the powerful role it plays.[20]

Example 5: Coding and Analyzing Archival Data— or, Secondary Analysis[21] (Chapters 5 and 6–8)

Archival data are collected and paid for by someone other than the researcher.[22] One of the most common types of archival research is the literature review. A meta-analysis is a literature review that results in a *quantitative* summary of research findings; this means that numbers predominate in coding and analysis. But the first step in a meta-analysis is a *qualitative* determination of the eligibility of studies for inclusion. And *graphic* techniques, such as funnel plots, are usually considered essential for discovering important patterns in the data and for depicting a summary of the findings of research articles. The qualitative assessments of eligibility are combined with graphic depictions of patterns and numerical statistical summaries of results to produce an overall summary. Another important field of research using archival data is the study of social media. Millions of messages can be gathered, coded, and analyzed quantitatively, qualitatively, and visually. Visualizing information is often indispensable for discovering comprehensible patterns in the huge amounts of data available from social media, as well as from other archival sources.

Example 6: Coding and Analyzing Data from Combined Designs (Chapter 13 and throughout)

Our general point in the first five sets of examples is that combined methods of coding and analysis are common in all designs, even those ostensibly tied to quantitative, qualitative, or graphic methods of analysis. A fortiori, if it is true of unitary designs, it will be even truer of explicitly combined/mixed designs. In combined designs, it is especially important to ensure that your coding methods are compatible. It is crucial that you do not assign incompatible coding schemes to data that you intend to merge for analysis. If you intend to unify your analysis only at the more theoretical and *interpretation* stages, then the coding for quantitative and qualitative *data* may remain distinct.

Here are two examples: Say you are investigating the quality and quantity of food available in low-income urban neighborhoods. Both quality and quantity are important attributes of food availability, but your coding must reflect the interaction of both attributes. Is a lot of poor-quality food better than a little high-quality food? Is quantity better measured by weight, volume, or calories? What attributes of food indicate "quality"? If your coding decisions skew your analysis, you might even conclude that a small amount of bad food is a good thing. Or to take a second example: Say you are trying to determine the adequacy of school facilities for the next 20 years. You use school-age population projections from census data. You determine population trends by county and then create qualitative categories, such as *rapidly increasing, increasing, stable, declining*, and *rapidly declining.* You might then create a color-coded map by county to

---

[20] Cross and Parker (2004); Castells (1996).

[21] For secondary analysis of "big data" from the Census Bureau, see Capps and Wright (2013).

[22] This definition comes from the classic discussion in Webb, Campbell, Schwartz, and Sechrest (1966).

determine regions of the state in which schools may not be able to house their students or in which school buildings may be empty in coming years. Where you place the "cut scores" to determine the category boundaries matters greatly; it could mean the difference between accurately or inaccurately determining school capacity and could greatly influence policy decisions affecting many people.[23]

LOOKING AHEAD

The preceding six sets of examples correspond to chapters in Part I on coding choices and are related to how those choices link to selecting analysis methods in the remaining chapters of the book. Each chapter on coding choices includes suggestions about which chapters to consult for analysis options (in Parts II and III). Those analysis chapters also discuss interpreting and reporting your analytic results by addressing the questions: How do you make sense of your analytic results, and how do you convey your interpretations to others? Although the coding and analysis sections of the book are closely related, there are important differences among them.

The chapters in Part I on coding are organized by design; each is relatively freestanding and can be read independently of the others. The chapters in Parts II and III, on methods of analysis, are more closely tied together. This is especially true of Part II, on quantitative analysis. The later chapters in Part II often assume knowledge of the earlier. Also, the analytic techniques in Part II are routinely used together in practice; researchers frequently use all of the types of quantitative methods—descriptive, inferential, and associational—in a single project. The inductive and deductive methods discussed in Part III, on the analysis of qualitative data, are less often employed together in a formal way. But they are often used together informally, perhaps even autonomically. Induction and deduction are, like inhaling and exhaling, ultimately inseparable, as are, we believe, qualitative and quantitative concepts.

Probably the most exciting moment in a research project occurs when the results from the data analysis start to become clear and you can actually begin to interpret the findings.

That is what it was all about. You've struggled devising a good research question, selected an appropriate design for gathering the data, identified a justifiable sample, and had it all approved by the institutional review board. And now, at last, you are going to see how it turned out. Will the painstaking and detailed work pay off? Your work is more likely to yield something interesting and important if you have given serious consideration to alternate methods of analysis. If you have done that, your choices were made knowing the options. It is hard to make a good decision otherwise. Our goal in this volume is helping with that penultimate, and crucial, step in a research project—choosing the most effective methods of data analysis.

[23]For an example of this type of population prediction being used in a policy context, see Simon (2012). For a discussion of how population predictions based on prior trends and assumptions may be misleading and therefore require adjustments in analysis methods, see Smith (1987).

CHAPTER 1

Coding Survey Data

In this chapter we:

- Present an example of pitfalls to avoid when constructing surveys.
- Discuss what methods to use to write an effective questionnaire, including:
  - Considerations when linking survey questions to research questions.
  - When to use questions from previous surveys.
  - When to use various question formats.
    - When should you use open-ended or forced-choice questions?
    - When should you use reverse coding?
    - How many points do you need in a scale?
    - When should respondents be given neutral response options?
  - When mode of administration (face-to-face, telephone, or self-administered) influences measurement.
  - Steps you can take to improve the quality of questions.
    - Checklists, focus groups, expert review, linking survey questions to research questions, cognitive interviews, pilot tests, and survey experiments.
- Review coding and measuring respondents' answers to questions.
  - When can you sum the answers to questions (or take an average) to make a composite scale?
  - When are the questions in your scales measuring the same thing?
  - When is the measurement on a summated scale interval and when is it ordinal?
- Indicate where in this book to find further analysis guidelines for surveys.

If you have decided that your research question can be best addressed through survey research, you have chosen to use the iconic method for quantitative social research. When people think of gathering quantitative social science data, survey research often comes to mind, and it is true that more quantitative social science data have been collected through survey research than in any other way. Survey research has been a social science success story in fields as diverse as network analysis and election prediction. That success and widespread use has led to much accumulated wisdom about how to code survey questions and answers. So you will have many standard resources on which to rely.

Coding for surveys usually focuses on quantitative data, and we do so in this chapter, even though survey researchers often collect qualitative data when they ask open-ended questions. For such qualitative survey data, see Chapters 2, 4, and 11. Methods for coding quantitative survey data are often parallel to methods used to code quantitative experimental data. Although the two can differ in important ways, it can be helpful to compare them, and we do so at several points throughout this chapter. For an overview, compare the Summary Table in this chapter and the one in Chapter 3 on coding experimental data.

As in many other research designs, coding and measurement in surveys falls naturally into two phases: first, before the data collection, as you write the questionnaire;[1] and second, after the data collection, as you sort and categorize the responses to prepare them for analysis. In the first phase you construct the questions and, with forced-choice questions, you precode the answers (coding occurs later with open-ended questions, of course). In the second phase you continue coding the answers to prepare them for analysis. Surveys differ from most other designs in that the work in the first phase is more extensive. The contrast with coding semistructured interview questions is particularly sharp: In interviews, almost all the coding is done after the data are collected; in surveys, nearly all of it is done before.

The first phase of coding in survey research focuses on what questions to ask respondents, how to format the questions, and how to code the answers so that they can be analyzed and interpreted. Addressing the content and format of the items in a survey you write includes taking steps to increase the chances that the response options are actually measuring what you want to measure. In other terms, the focus in the first phase of coding in survey research is the **content validity** of the questions. The questions have content validity to the extent that they address your research questions and the theoretical substratum on which they are built.

In some disciplines, especially psychology and related fields, it is probably more common to use a preexisting survey than to construct one. For example, some 3,000 commercially available instruments are indexed in *Tests in Print* and reviewed in its companion volume, the *Mental Measurements Yearbook*. And many more, such as *General Social Survey*, are essentially in the public domain or are available from individual scholars at no cost. We begin by discussing constructing a survey rather than reusing one.

AN EXAMPLE: PITFALLS WHEN CONSTRUCTING A SURVEY

It is crucial to do all that you can to write a good questionnaire. No one would disagree, but sometimes researchers appear to forget that the answers to survey items can only be as good as the questions. You can't get good answers unless you ask good questions.

[1]We use the term *questionnaire* in the generic sense of any standardized list of questions, not only in the strict sense of a list that respondents read and answer in writing.

Writing survey questions requires a great deal of thought, skill, and attention to detail. It is something that almost no one working alone and writing only one draft can do well. As consultants helping people with their research, we have seen a remarkable number of nearly useless surveys. A consultant is usually brought in to help in the analysis phase, that is, after the survey is written, administered, and the responses collected. This is too late. What you should do before it is too late—before the analysis phase—is the focus of this chapter. The following example of a survey we were asked to help with (details are disguised to ensure anonymity) illustrates some pitfalls of putting off crucial work until it is too late.

The survey was self-administered. Instructions were ambiguous for some questions.[2] Because these questions could be, and actually were, interpreted in more than one way, the same was true of the answers. There was nothing to be done but to discard those questions. For some other questions, the response categories were not appropriate. One question, for example, did not have a "does not apply" option, although the question clearly did not apply to a large number of the respondents. On some other questions, the responses either were not mutually exclusive or were not exhaustive. **Mutually exclusive** answers contain no possible overlap; more than one answer cannot logically be chosen. For example, "How old are you: (a) 20–30, (b) 30–40, (c) 40–50, (d) 50–60?" does not provide mutually exclusive answers; someone who is 30 could answer either (a) or (b). **Exhaustive** answers cover all possible options. "Are you Protestant or Catholic?" is not exhaustive; it leaves out, among others, Jews, Moslems, Hindus, and the nonreligious. Mutually exclusive and exhaustive are perhaps the two best-known criteria for question options, as well as for any system of categorization, but it is remarkably easy to slip up and write a poor question with answers that are not exhaustive and/or mutually exclusive.

Of the 65 questions on the survey we were helping with, 40 seemed to have no major problems. After pointing this out to the client, he said, "Okay, could you help me analyze those 40?" Our reply was: "Sure, what do you want to know?" He stared at us blankly. He could not easily articulate the research questions that he hoped to answer with the responses to his survey questions. He thought that his research questions were implicit in his survey questions. We explained as gently as possible that with 40 questions, he had 760 bivariate relationships that he could examine and many more multivariate relationships. For example, Question 3 asked about respondents' education levels. Question 11 asked about political party identification, and Question 21 asked about attitudes toward a new law. Did he want to know about education's relationship to attitudes, or party identification's relationship to attitudes, or the effect of education on attitudes controlling for party identification—or what? Eventually, we were able to work with him to construct some research questions—questions that had been in the back of his mind when he wrote the survey. We could then relate these research questions to some of the survey questions, but the process was frustratingly inefficient for the client.

Because the survey author did not attend to the first phase of coding and measurement, he had largely wasted his time. Even worse, he had wasted the time of the hundreds of survey respondents who answered his survey questions. And, had the survey been better constructed, our time would also have been spent more effectively; but,

[2]See the later discussion of steps to improve the quality of questions.

unlike the survey respondents, at least we got paid for our efforts. The survey author was very intelligent, but he was ignorant of or did not pay sufficient attention to some basic procedures for writing survey questions. In survey research, the "up-front" work is exceptionally important, more so than most people realize. It determines everything else.

Numerous excellent books are available to guide researchers in writing questionnaires. It would be foolish not to spend at least a few days consulting such works on question design before drafting your survey.[3] In the following pages, we outline some of the steps to take in order to avoid disasters such as those in the preceding example. However, the treatment here is necessarily brief, so once you have settled on your general approaches, you will want to consult more specialized and detailed works; those we have cited are meant to suggest places to start. A book such as ours is like a regional map for someone traveling to a city. It is good for orienting yourself and getting there, but to find your way around the city once you have arrived, more detailed maps are needed.

WHAT METHODS TO USE TO CONSTRUCT AN EFFECTIVE QUESTIONNAIRE

Considerations When Linking Survey Questions to Research Questions

First, if you have decided to use a survey design, you should have already judged that potential respondents are likely to have knowledge sufficient to answer your questions or that they have beliefs that are clear enough to respond meaningfully. For example, if you asked us whether we favored proposed tariff regulations concerning the import of mineral ores, we wouldn't have enough knowledge to have a belief. We might be able to offer an opinion based on vague attitudes about tariffs, but is this what you would want to know?

Second, if you have chosen to conduct survey research, you have thereby already decided to use mostly structured questions designed to yield structured answers. Questions that might be just the ticket for an interview—such as, "What's it like living around here?"—would not work well on a survey. A survey question on the same topic might take the form, "In comparison with other neighborhoods where you have lived, would you say this one is safer, less safe, or about the same?" Survey questions are written to produce easily codable responses, and you should have already decided when you settled on a survey design that your research questions could be answered, for the most part, by short, structured responses. If the questions cannot be so answered, then survey research is probably the wrong design for your purposes.

There are two broad categories of surveys and survey questions: those that ask for facts and those that ask for attitudes, beliefs, or opinions. The kind of questionnaire you write will depend on your research question(s). These two types of survey questions

[3]Two classics are Sudman and Bradburn (1982) and Fowler (1995). A more recent and comprehensive treatment is Presser et al. (2004). See also Fowler (2008).

collect what can be called either **objective data** or **subjective data**.[4] The terms *objective* and *subjective* can be hard to define and are often quite controversial, but in the context of survey research, there is a clear distinction. If you can only reasonably get data from the *subjects* of the research (respondents to the survey), then the data are subjective. Opinions are a good example. If it is possible to get data other than from the respondents, such as their place of residence or age, then the data are objective. To answer your research questions, do you need objective or subjective data? Most often, perhaps, you will need both, and your survey questionnaire will seek to gather both kinds of data. Even when the focus is on subjective data, such as beliefs and attitudes, factual, objective information is usually collected as well, such as respondents' ages, genders, education levels, incomes, and so on.

Coding issues can differ for objective and subjective survey data. For objective data, knowledge and memory can be big issues. For subjective data, they hardly ever are. Coding usually is not much of a problem with objective data, assuming respondents know or remember what you ask them: How many times did you visit the dentist last year? When did you first become employed full time? The answers are easily coded, but it would also be easy for respondents to forget the exact details. To help respondents, it is often useful to ask them about a range of values rather than exact values—for example, *not at all, once or twice*, . . . *more than 5 times*.

By comparison, memory and knowledge are not as often a problem when survey researchers ask respondents for subjective data. It is usually safer to assume that people know how they feel, and surveys rarely ask them to remember how they felt in the past. On the other hand, coding and interpreting answers to questions seeking subjective data can be very complicated. In brief, for objective data, knowledge and memory can be problems, but coding is usually easy. For subjective data, coding is usually hard, but the knowledge and memories of respondents usually are not at issue.

How do you link your survey questions to your research questions? It would hardly ever be appropriate simply to turn your research question into a survey question. Say that your research question is, What is the relationship in this population between age, education level, and income? It would probably not be productive to ask respondents an open-ended question about this. Rather, you would ask short factual questions of respondents and make inferences yourself. Even when seeking answers to questions about subjective matters, such as respondents' feelings, it is rare to have respondents speculate on this. The research question might be, What is the relation between job satisfaction on the one hand and feelings of anxiety on the other? A good research question is almost never a good survey question for respondents. To answer the research question about job satisfaction and anxiety, you might ask as many as a dozen questions, half of them to measure job satisfaction and half to measure anxiety. The dozen questions would constitute your operationalizations of the variables *satisfaction* and *anxiety*.

To move from your research questions to your *first draft* of the survey questions, the first step is to determine what your variables are. These should either be explicit or

[4]Some researchers object to the use of these terms, mostly because the distinction between them can be drawn more sharply than is warranted. But we find the concepts useful and can think of no equally good labels. The concepts and labels are much debated; for enlightening discussions, see Hammersley (2011); Letherby, Scott, and Williams (2013); and Daston and Galison (2007).

implicit in your research questions. Make a complete list of these variables. Put them into one column of a two-column table. In the second column, write a draft of the question or questions you will use to gather data on each variable, or find questions from previous surveys to use to answer your research questions.

When to Use Questions from Previous Surveys

It is rare to plan to do survey research on variables that have never been studied before. Because writing your own questions is hard work, and work with many pitfalls, it is crucial to review the literature in your field before deciding to compose your own questions. Think hard about how your variables have been coded and measured by others. Literature reviews yield precious information not only about substantive findings but also about methodological procedures, such as ideas about how to construct your measurement instrument.

Variables such as anxiety and job satisfaction have been studied and measured by many researchers. Give serious consideration to asking your respondents questions used in previous studies. There is a strong presumption in favor of using instruments (in whole or in part) developed by other researchers. If they have done good measurement work on their questions, this can save you an enormous amount of time. You will still have to do some of this work with your data, such as examining the reliability of any scales as they were answered by your respondents. Here is a *very important* point: You should not use others' questions as an excuse to skip testing for reliability. Rather, you use others' questions because they serve as a sort of pilot test for your survey and because it is very helpful to be able to compare your results with previous work.

Another advantage to using existing measurements is that doing so facilitates the study of change. Perhaps you want to investigate whether the relation between job satisfaction and anxiety changes with economic conditions. If you have an effect size for the relationship based on data collected during a previous recession, you can compare that with the effect size of the relationship in more prosperous times or in a current recession. You complicate your work greatly if you use a different measure. As the old saying goes, "If you want to measure change, don't change the measure."

However, just because there is a preexisting measure of a variable, it does not necessarily follow that it is appropriate for your purposes and that you should use it. One of your important hypotheses might be, "The reason researchers studying my topic have gone astray is that they have used poor measurements." Although there are many benefits to reusing items, don't be afraid to revise another researcher's questions. However, there is much to be gained from using the same questions. In brief, one rarely has to and rarely should start from scratch when writing survey questions. Much can be learned through replication and through modification and reanalysis of responses to existing survey questions.[5] Of course, if you use others' questions, you will need to obtain permission (from the author and/or the publisher) to do so, perhaps paying a fee, and to cite their work as appropriate.

[5]Replication can be tricky; accuracy depends on exercising extreme care. See Altman and McDonald (2003).

When to Use Various Question Formats

The range of possible formats for questions is wide. The first division is open-ended versus forced-choice questions. Do you want your respondents to answer questions freely in their own words, or do you want them to select among a set of predetermined options? Say that you are asking questions of clients of the Job Outreach Bureau (JOB) in order to evaluate that office. An open-ended question might be as follows:

1. Please tell me, in your own words, what you think of the Job Outreach Bureau (JOB). Write on the back of the page if you need more space.

Questions such as this one have much to recommend them, but you will want to make limited use of such open-ended questions on a survey, for two main reasons. One has to do with measurement; the other concerns resources. First, respondents tend to skip such questions, and that raises problems of response bias and missing data. Second, open-ended questions take many more resources than forced-choice questions to code and analyze. Just as survey respondents tend to think it is too much work to answer open-ended questions, you may think it is too much work to code and analyze the answers to them. You may have decided to do a survey because your research question requires responses from a large, representative sample. If so, you probably have already determined that in order to code and analyze answers from hundreds of respondents to dozens of questions, you have to use forced-choice questions almost exclusively. You are willing to pay a price for that. You lose the depth and nuance possible with open-ended questions. But you gain the breadth and generalizability possible with a large sample survey.

If you decide you need to use a forced-choice question to obtain a general evaluation of the JOB, you could give respondents a rating scale such as the following:

2. On a scale of 1 to 10, with 10 being high or positive, how would you rate JOB?

   1 2 3 4 5 6 7 8 9 10

   (Please circle the number that best expresses your opinion.)

In an ideal world in which both you and your respondents had a great deal of time, you might want to ask both Question 1 and Question 2. Comparing the answers to the two questions could be very informative. If you can draw the same conclusions from the two types of questions, this provides cross-validation. You can be more confident about what you have learned than if you had used only one of the questions. Or, what you learn in the open-ended paragraph could help you explain the answers to the forced-choice rating scale. It is also possible that respondents will give conflicting answers to the two questions, perhaps giving a good "grade" on the rating scale but complaining about the inadequacies of the bureau in the open-ended answer. If your resources allow you to gather your evidence in more than one way, you are better off having both types of questions on your survey. You could also make your choice about type of question not on substantive grounds about what you'd like to know but because you believe that respondents would be more likely to answer some kinds of questions than others. Combining broad statistical approaches with in-depth methods that yield qualitative data is a good idea whenever you have the resources. Teams of researchers working on projects may have resources sufficient to do extensive multimeasurement work. Solo investigators can rarely afford to do a great deal of it. But they can do more than was once the case because of the wide availability of computer software packages for textual analyses and the increased possibilities for combining the analysis of qualitative and quantitative data.[6]

Surely the most common question format in survey research today is the **Likert scale**, named after Rensis Likert, the investigator who pioneered it. Respondents are given a series of statements with which they agree or disagree. The familiar set of choices is a 5- or 7-point scale that ranges from *strongly agree* through *neutral* to *strongly disagree*. Returning to our example, the clients of the JOB could be asked to agree or disagree with the statements in items 3 through 6, as follows:

3. JOB found opportunities for me that I wouldn't have been able to find on my own.

   Strongly agree Agree Neutral Disagree Strongly disagree

4. JOB increased my self-confidence in employment interviews.

   Strongly agree Agree Neutral Disagree Strongly disagree

5. JOB was less helpful than I expected it to be.

   Strongly agree Agree Neutral Disagree Strongly disagree

6. It would be better to replace JOB referrals with an actual training program.

   Strongly agree Agree Neutral Disagree Strongly disagree

This format is widely used because it has many positive features. Scores on the questions can be summed to get an overall assessment of the JOB office, but each question's score provides specific information about an aspect of the office.[7] This is more informative than an overall rating scale alone. One can test sets of questions for reliability, and one can learn from the scale's component questions why respondents tend to rate it as they did. Perhaps those who gave the JOB high rankings were especially fond of the way it helped them with their self-confidence and those who gave it low ratings did so because they thought it should focus on training, not on referrals. It is also helpful that the statements can be positive (such as Questions 3 and 4) or negative (Questions 5 and 6). This enables you to avoid the kind of bias that might occur if respondents liked to agree or to disagree with whatever is said to them. Some researchers think yea-saying or nay-saying is a big problem. It should be headed off by wording some questions positively and others negatively. (Of course, when you do so, you will code the negative question responses on a reverse scale, as we discuss subsequently.)

The Likert-type format is also preferable to the one that asks respondents to check all of the options that apply. Evidence from survey experiments very powerfully

[6]For an example of the very fruitful integration of survey and interview data, see Pearce (2002). For an account of the possibilities of software approaches to uniting text and numerical data, see Bazeley (2006).

[7]Likert scales have a long history of use in the social sciences; see Spector (1992). Typically the means of scales are analyzed, but more advanced options using other characteristics of the responses (such as their skewness and kurtosis) can also be very revealing; see Camparo and Camparo (2013).

demonstrates the superiority of the forced-choice format, in which respondents have to answer a question about each part of the topic. By contrast, the check-all-that-apply format is less likely to encourage respondents to take the questions as seriously or to think about their responses carefully.[8]

When Should You Use Reverse Coding?

This is one of several measurement issues with a set of questions, such as the four on the JOB office, that involve the wording of questions and the assigning of numbers to the answers (coding). It is often advisable to use both negatively and positively worded items, as there are some grounds for worry that some respondents like to check "agree" just to be agreeable—or to disagree to be disagreeable.[9] If you do use positively and negatively worded items, you will need to use reverse coding before summing the scores on items to make a scale. For example, for Question 3 you might give a 5 to *strongly agree*, a 4 to *agree*, and so on. In this case, for Question 6 you would give a 1 to *strongly agree*, a 2 to *agree*, and so on. This reverse coding is required because someone who says "strongly agree" to Question 3 likes the JOB office, but someone who says "strongly agree" to Question 6 dislikes it. One of the most common sources of puzzling results, such as an item that is highly inconsistent with others in the scale, is the investigator's having forgotten to reverse code an item. This is a rookie mistake, but one that is also made surprisingly often by veterans.

How Many Points Do You Need in a Scale?

The short answer to the question of how many points or options you should provide is that you should provide as many as are meaningful. If you err, it is better to include too many than too few. Neither is ideal, but you can always combine answers if you have too many options. Of course, after the survey, it is impossible to expand the number of options if you have too few. For some simple questions, *agree–disagree–unsure* may be enough. For others, on which you think respondents might have many levels of feeling or opinions, a scale with as many as nine *agree–disagree* intervals might be appropriate. Some researchers use a "feeling thermometer," on which respondents can pick a point on a thermometer ranging from 0 to 100. Commonly on Likert scales, one uses a range either of 5 or 7 points. An odd number of points on the scale makes it possible to have a neutral response choice in the middle.

When Should Respondents Be Given Neutral Response Choices?

Should the questions include neutral responses such as "don't know" or "unsure"? Measurement specialists disagree about this. Many researchers recommend eliminating all neutral, wishy-washy options. They argue that by forcing respondents to take a stand, you get better answers, that is, scores with bigger variances. Although it is true that you get bigger variances this way, that alone does not justify eliminating the neutral option

[8] Smyth, Dillman, Christian, and Stern (2006).

[9] One classic study found that such yea-saying and nay-saying differed by respondents' race (Bachman & O'Malley, 1984).

in all questions. For some questions, "don't know" or "don't care" are real opinions worthy of investigation. If you use a forced-choice format, the choices you force respondents to make should be good ones. In addition to being exhaustive and mutually exclusive, the choices have to be *valid*, which means that the question options should capture what the respondents actually believe. Forcing respondents to act as though they know or they care, even when they do not, reduces validity. Researchers may find it inconvenient if respondents frequently pick the neutral response; it tends to reduce variances and make it more difficult for researchers to get statistically significant results. But that is the researchers' problem, not the respondents' problem.

When Does Mode of Administration (Face-to-Face, Telephone, and Self-Administered) Influence Measurement?

Advantages and disadvantages of various modes of surveying are reviewed in our companion volume on choosing a research design.[10] Here we briefly mention some of the more frequent coding and measurement problems that are associated with particular modes of survey administration.

Face-to-face surveying raises the issue of the gender or color or age of the faces. There is no doubt that at least some respondents react to the characteristics of the survey interviewer, as well as to the questions. To compensate for this, you could randomly assign survey interviewers to respondents, which could randomly distribute such biases. You can also investigate whether responses vary with the characteristics of interviewer and respondent. It is often possible to check to see whether responses vary according to the age, gender, and ethnicities of respondents and survey interviewers. If they do, you can statistically control for these variations. Of course, every control variable you add means that you will need to increase your sample size. And there are other interviewer characteristics that could be considered, such as accent and mode of dress.

Telephone surveys—we might call them "ear-to-ear"—are very common for obvious reasons: They greatly reduce the time and cost of surveying large numbers of respondents spread over a wide geographical area. Some respondents will find the telephone intrusive, but others will prefer it or will only be willing to be contacted by telephone.[11] And, like the face-to-face survey, the telephone researcher is available to clarify survey questions that the respondent does not understand. Respondents may still react to gender, ethnicity, and tone of voice differences in telephone surveys, so it remains important to statistically control for any such biases in responses.

One of the important advantages of **self-administered** surveys is the elimination of this kind of bias. You should try to design self-administered questionnaires so that it will be *extremely* difficult to misinterpret the instructions. But some respondents will almost certainly misinterpret them. Never underestimate the inattention of respondents. They will almost never find your survey as interesting as you do. Probably the most frequent form of misinterpretation of survey instructions occurs when there are **skip patterns**: "If you answered yes to Question 14, go on to the next question; if you answered no, skip ahead to Question 18." A remarkable number of people can be inconsistent when faced with this kind of question. They will answer "no" to Question 14, forget to skip

10 Vogt et al. (2012, Ch. 1, pp. 19–23).

11 Stephens (2007).

ahead, and answer Questions 15, 16, and 17 as if they had answered "yes" to Question 14. For example, Question 14 might be, "Did you work while in college?" Questions 15, 16, and 17 might be about how much you worked and the nature of the work. If you answered: "No, I didn't work," then questions about the number of hours worked and the nature of the work are inapplicable, but sometimes people will answer them anyway. Such logically inconsistent responses are generally not usable, and you have to discard the data—never a pleasant experience.

If you must use skip patterns, consider face-to-face or telephone survey administration in which the interviewer does the skipping. Another alternative is self-administered electronic Web surveys; these can be designed so that the program does the skipping. That way the respondents never get a chance to answer questions they should have skipped.

The key point to remember is that survey experiments have shown that very small differences in question format can produce big differences in results, sometimes bigger even than those produced by differences in the content of the questions. Reviewing the results of such research on survey research is always time well spent.[12] A classic example of the influence of wording is: "Should the government not allow *X*?" versus "Should the government forbid *X*?" Although not allowing and forbidding seem logically equivalent, many more people will agree that the government should not allow something than that it should forbid it. The two apparently have different connotations for many people.

What Steps Can You Take to Improve the Quality of Questions?

After you have written your initial draft of questions based on your literature review and on your research questions, there are several steps you can take to improve the quality of your questionnaire. By *quality*, we refer here to the meaning of the questions, not to their technical aspects, such as when you need to use reverse coding. Constructing a good survey—one in which the questions are valid and truly ask about what you want to know—is a difficult process that requires many steps.[13] Most solo researchers will not be able to do all of them that we list, but doing them all should be kept in mind as an ideal. After each step you make the necessary revisions and proceed to the next step. In order, the steps are:

1. Review your draft using a checklist designed for the purpose.
2. Conduct focus groups with people who would qualify as potential respondents to help you make sure you have not omitted important items.
3. Have a small panel of experts review your questions; ideally, they would have expertise in question design, as well as in the topics being studied.
4. Review your revised questions in terms of your research question and your analysis plan, specifically, how you will be able to tie the responses to your research questions.

---

[12] A good example is Christian, Dillman, and Smyth (2007).

[13] See Sanders et al. (2010) for a discussion of the numerous ways survey respondents interpreted the phrase "had sex" and the implications of multiple meanings for misclassification bias.

5. Interview people who would qualify as potential respondents to ask them about the content and quality of the questions.
6. Pilot-test the survey with a sample of real respondents.
7. For questions that remain unclear, conduct survey experiments.

1. **Review your draft survey using a checklist.** This is the minimum first step. Writing surveys is a complicated business, and it is easy to omit something important or to make an easily corrected mistake. An excellent checklist is the one by Gordon Willis.[14] We have worked on several surveys over the years and still find it useful, as have many of our students. There are too many elements to the process of constructing survey questions—both their format and content—to trust things to memory. Even pilots who have flown planes for decades do not skip their checklists—or they are foolishly (and maybe criminally) negligent if they do.

2. **Conduct focus groups with potential respondents.** Focus groups sometimes come up with insights that the same people answering questions individually do not. The idea is to get a group of similar people (their similarity will be that they are the sorts of folks you will sample for your survey) to focus on something, in this case the scope of your questionnaire. Focus groups are a good place to ask and learn about problems with the overall presentation of the survey—clarity of instructions, length, question order, and so on.

3. **Have a small panel of experts review your questions.** Anybody is better than nobody. But it is nice to have someone who knows about the subject and somebody (it could be the same person, of course) who knows about survey design. Doctoral students have a ready-made panel—the members of their dissertation committees. Three or four experts are usually enough, but if you can importune half a dozen or more, that would usually be an advantage. Even when you get contradictory advice, you can use it to stimulate your thinking about your survey. And you will almost certainly get contradictory advice—about reverse coding, neutral options, question wording, question order, and so on.

4. **Review your revised questions in terms of your research questions and analysis plan.** Make sure that you have a sufficient number of questions for each of the variables and concepts contained in your research questions and that the questions will be such that you can use the answers to address your questions in the analysis phase. For example, if your dependent variable is measured with one yes–no question, you will not be able to use ordinary regression analysis to interpret it. In general, scales are better than individual questions. The more important the variable and the more difficult it is to measure, the more important it is to use multiple measurements of the variable, that is, multiple questions. Generally best practice for turning the multiple measures into a scale is to use structural equation modeling, because it allows you to construct a continuous latent variable out of categorical indicators (see Chapter 9).

[14] The checklist is available in Willis (2005); for an online discussion see: *http://appliedresearch.cancer.gov/areas/cognitive/interview.pdf*. For further discussion in the context of cognitive interviewing (Step 5), see Beatty and Willis (2007).

5. **Interview potential respondents to ask them about the content and quality of the questions.** This is often referred to as **cognitive interviewing**, which is usually more targeted than interviews with focus groups, which tend to be done at an earlier, more exploratory, stage of the question writing. Cognitive interviewing focuses on whether survey questions are eliciting from respondents the kind of information that the researcher means to elicit. In a word, cognitive interviewing is about the validity of questions. Evidence about validity is obtained by asking a small sample of respondents to tell survey interviewers what they meant and what they were thinking as they answered questions. The researcher uses what is learned from their responses to revise questions.

6. **Pilot-test the survey with a sample of real respondents.** We have never successfully anticipated every problem with a survey, but we have come closest after having conducted serious pilot testing. If you are developing your own scales, you *need* quite substantial pilot testing. Indeed, scale development is itself a discipline. Not infrequently, a researcher trying to learn about a topic has to develop an instrument to study it, and the instrument can be an important contribution to knowledge in its own right. Sometimes a good instrument has made a more lasting contribution to research than the findings of the research project on which it was first used.

7. **For questions that remain unclear, conduct survey experiments.**[15] It is often the case that issues and uncertainties remain even after the previous steps have been taken. Here is where **survey experiments** become very helpful. For example, to study the effects of different question wordings or question orders in your survey experiment, you would randomly assign members of the sample to different question wordings or question orders and test the effects of these differences. If there is no difference in response patterns between the different wordings or orders, combine them and analyze them together. If there is a difference, you have learned something psychometrically interesting. One of your findings will be how different wordings made a difference. Report any differences. Finally, the survey experiment has one special strength: It is one of the few research designs that commonly attains the ideal of combining random sampling with random assignment. The special strength of experiments is internal validity achieved through random assignment. The special strength of surveys is external validity achieved through random sampling. A survey experiment can unite the two in one study.

CODING AND MEASURING RESPONDENTS' ANSWERS TO THE QUESTIONS

After you have produced the best survey you can using guidelines such as those just discussed, and after you have received the responses, then what do you do? As Phase 1 tied the survey back to the research questions and was mostly related to validity, Phase 2 looks forward to the analysis and is mostly related to reliability.[16]

[15] Gaines, Kuklinski, and Quirk (2007) are superb on survey experiments.

[16] For a general discussion of validity and reliability in data coding, see Vogt et al. (2012, pp. 317–333).

Your first step is to enter the responses into a spreadsheet or a statistical package. Each row is a respondent; each column is a variable or question. This is a very straightforward process. Open-ended questions are somewhat more complicated, and you have more choices. The answers to the questions could simply be entered into word processing software to prepare them for analysis. Software packages for text data are also useful for the purposes of integrating qualitative and quantitative data analysis.[17] On the other hand, if the answers to the open-ended questions are fairly short, they can be entered directly into a spreadsheet or statistical package. Of course, if you have coded open-ended questions into briefer codes, probably as categorical or rank-order variables, you will usually want to enter those codes into a statistical package so that you can link the numerically coded to the categorically coded answers.

When Can You Sum the Answers to Questions (or Take an Average of Them) to Make a Composite Scale?

Summing the answers to questions with Likert scale responses (*strongly agree, agree* . . .) is the usual practice. Returning to our questions about the Job Outreach Bureau (JOB), if you added together the numerical codes of the answers, you could construct a "summated scale." The highest possible score would be 20 (5 times 4 questions = 20), and the lowest possible score would be 4 (1 times 4 questions). The advantages of using a scale, rather than studying individual items one at a time, are considerable. The meaning and interpretation of any one question are uncertain. Scales are generally more reliable and valid. It is easy to see why. Think of a multiple-choice examination on your knowledge of a subject. One question, no matter how good, would almost certainly be a poor measure of your knowledge. To measure what you really know about the subject, a fairly large number of questions would be needed, and, within reasonable limits, the more questions, the better. The same is true of respondents' beliefs and attitudes on complex matters. One question will hardly ever do the job well. When you use a scale, you should test for reliability using a technique such as Cronbach's alpha (see the next subsection) or factor analysis.

On the other hand, when you sum a group of items into a scale, you may lose important information about differences in individuals' scores. For example, on the four questions in the preceding illustration, someone who answered neutral on all four questions would get a total score of 12 ($4 \times 3 = 12$), whereas another respondent who answered "strongly agree" on two of the four and "strongly disagree" on the other two would also get a score of 12 ($5 + 5 + 1 + 1 = 12$). But the two response patterns are *dramatically* different—as different as they could possibly be. The first respondent has given uniformly neutral answers, whereas the second has given sharply discordant responses. This is the reason that, when summing items in a scale, you should also explore patterns in the responses with exploratory data analysis (see Chapter 6).

Also, although there is general consensus that scales are better than single items, the common practice of summing Likert scale items is more than a little controversial. Can Likert scale items correctly be treated as interval-level data, in which case they can correctly be summed, or should they be treated as ordinal, in which case they cannot?

[17] One popular program is NVivo; see Bazeley (2007).

Although the practice of scaling ordinal variables is widespread, it has been challenged. It is hard to justify the assumption that the distance between, for example, *strongly disagree* and *disagree* is the same as the distance between *disagree* and *neutral*. If the distances between the points on a Likert scale are not equal, then summing items is at best dubious, though widely practiced by applied researchers.[18]

When Are the Questions in Your Scales Measuring the Same Thing?

Because of the increased reliability and validity that can result from multiple measures of a variable, it is generally advisable to write several questions, the answers to which you expect to combine into a more general measure. But how do you know whether you have been successful in writing questions that consistently measure aspects of the same variable? This is a question of **reliability**. The most common measure of reliability for survey scales is **Cronbach's alpha**. Cronbach's alpha is a correlational measure of the consistency of the answers to items in a scale. For example, it would tell you the extent to which people who answered some questions favorably or unfavorably about the JOB tended to do so on all of the items. If they did not, then it is probably the case that the items are not really measuring aspects of the same thing, and the items should not be summed up to make an overall rating scale. If your items seem not to be measuring aspects of the same thing, don't discard them. They may be measuring two or more distinct things. And you can always analyze the answers to the individual questions, even when adding their scores together would be inappropriate. Your intention to write questions that can be summed into a scale is no guarantee that the questions will work as you intended. You cannot know in advance whether your efforts to write questions that are related have succeeded. They may be related in your mind, but are they related in your respondents' minds? To find out you have to probe their minds by collecting and analyzing their answers to the questions.[19]

Cronbach's alpha ranges from 0 to 1.0—from answers that are completely unrelated to those that predict one another perfectly. A common threshold for scale reliability in survey research is .70. An alpha that high or higher is evidence that the questions on the scale are measuring the same underlying concept.[20] Quite often, as you read survey research reports, you will see much lower reliabilities. Any conclusions drawn from scales with reliabilities lower than .70 should be treated with extreme skepticism. A reliability score of .70 is the *minimum* acceptable. For example, imagine that in the real world, one variable *completely* determines a second variable. But, if your measures of the two variables have reliabilities of .70 and .60, the highest possible correlation between the two would be .65. The r^2, or coefficient of determination—the extent to which you can predict one variable using the other—would be .42—and this for variables that should be, if accurately measured, perfectly correlated: $r = 1.0$. In short,

---

[18] One classic discussion is Duncan and Stenbeck (1987). If you want to use multiple Likert scale items, many scholars would advise that it is better to use the items as multiple indicators of a latent variable in a structural equation model (see Chapter 9).

[19] For a summary of reliability measures, including Cronbach's alpha, as they are used for experimental data, see Chapter 3, Table 3.1, page 82 in this volume.

[20] For a good and highly cited overview, see Cortina (1993); for a more advanced discussion, consult Sijtsma (2009).

the more important the variables and their accurate measurement are to your research questions, the higher the reliability you should try to attain.

How do you tell whether your set of questions, which you intended to be one scale, is actually two or more scales? Factor analysis can be used for this purpose (see Chapter 9). Factor analysis is most appropriate for longer scales with numerous questions, and it requires a substantial sample size. Although it is more complicated than Cronbach's alpha, the principle behind the two measures (and other measures of reliability) is the same. And the reason for their importance is the same: Measures of reliability are crucial for validity, because a completely unreliable scale measures nothing and therefore cannot be a valid measure. Both exploratory and confirmatory factor analysis can have an important place in assessing and improving the coding and analysis of your survey.[21]

When Is the Measurement on a Summated Scale Interval and When Is It Rank Order?

The distinction between a rank order scale (very high, high, medium, low, and very low) and an interval scale (5, 4, 3, 2, and 1) can be murky. When you add together the rank order answers from several questions, does that transform them into interval answers? Not really, but the answers are treated by many researchers as if they were an interval scale. Strictly speaking, you should treat the scale of 4–20 from the above-discussed JOB survey as a rank order scale and use the appropriate rank order statistics.[22] On the other hand, the 20-point scale seems like more than a mere rank order scale. Many measures of many variables in the social sciences are treated as continuous, interval-level scales but might more correctly be thought of as "approximately interval." A rating scale composed of several Likert-like questions that has a high Cronbach's alpha can reasonably be treated in your analyses as a true interval-level measure even though it does not quite measure up, so to speak. In such cases we advise computing statistics using both the ordinal-level and the interval-level statistics (it only takes a few extra mouse clicks). Report and interpret *both* the ordinal and interval results—not just the one that conforms better to your biases.[23]

CONCLUSION: WHERE TO FIND ANALYSIS GUIDELINES FOR SURVEYS IN THIS BOOK

Writing good survey questions and effectively coding the responses is more difficult than many beginning researchers believe. Surveys are so pervasive in modern society that it is hard for people not to think that they know more than they do about what makes a good survey. On the other hand, some of the basic steps for increasing the quality of your survey are fairly clear, and we have reviewed those in the previous pages. They are also reviewed in the Summary Table on page 39. The more advanced steps for writing valid survey questions in such a way that you can code, analyze, and interpret

[21] Brown (2006).

[22] For example, Spearman's rho rather than Pearson's *r*; see Chapter 8.

[23] Spector (1992) is a good introduction.

the answers meaningfully are quite demanding. See the last paragraph in the Suggestions for Further Reading at the end of this chapter for some key sources.

Now that we have discussed how to code your survey data, it is time to make some suggestions about how to analyze the data thus coded. Virtually *all* analysis techniques can be used, and have been used, to analyze and interpret survey data. If you ask open-ended questions—and for your most important variables, we think this is a good idea whenever practicable—then you can use one or more of the techniques for coding and analyzing textual data. Both qualitative and quantitative approaches to the study of texts are available and are discussed herein, along with several software options (see Chapters 2, 4, 11, and 12). For analyzing answers to the more typical forced-choice survey questions, the most widely used techniques are associational methods based on correlation and especially regression (see Chapters 8 and 9 in this volume). Multiple regression is particularly suited to survey research, which usually gathers data about multiple variables and which often collects these data from samples large enough that multivariate analyses are possible.

Descriptive and exploratory methods of data analysis (Chapter 6) always have an important role to play when investigating any data, including survey data. And inferential, or hypothesis testing, techniques (Chapter 7) are routinely applied to most calculations of correlation and regression-based statistics. If your time is limited and you want to go to the most applicable (*not* the *only* applicable) analysis options, Chapters 8 and 9 are likely to be most useful for many readers planning to analyze survey data. In those chapters you will find details for specific types of correlations and regression analyses to use when investigating data coded in specific ways. If you are really short of time, you can get a quick overview of your data analysis options by consulting the Summary Tables in Chapters 8 and 9. Although the tables are integrated within and discussed in the texts of the chapters, we have done our best to make them freestanding.

The up-front work in survey research is very demanding, and, as compared with other forms of research, you often have to wait a long time until you begin to see some results or have any idea about how it will all turn out. Except for some hints you get during pilot surveys and other activities linked to writing the questions, the survey responses and preliminary data usually come in a rush. And the results of the first analyses can also come all at once. Thus surveys can require huge amounts of work with little payoff until the end, but the rewards at the end can be momentous. Of course, the quality of the answers you get from your survey questions is directly determined by the quality of the work you have done in preparatory stages described in this chapter.

Survey development occurs in two phases: constructing an effective questionnaire and coding respondents' answers to the survey questions. The Summary Table illustrates key considerations in each phase.

SUGGESTIONS FOR FURTHER READING

If you search the Internet for guidelines to coding your survey data, you will most likely find the Web pages of several companies offering to do this work for you—for a fee, of course. There are many reasons we do *not* generally recommend using such products. If you are a student, doing so may constitute academic dishonesty. Check the rules of your institution before spending the money. But the main reason we advise against using one of these services is that doing so severs the mental link between writing a survey and interpreting the results. It would be like planning a long hike, say on the Adirondack Trail, with the aim of writing a book reflecting on your experiences. But, rather than going on the hike, you hire someone to follow your planned route, keep a diary, and make extensive video recordings. Writing a book in this way would certainly be easier. However, most potential readers might not be interested if they knew that you did not actually experience the hiking described.

Many guidelines to writing, coding, analyzing, and interpreting survey data exist. In addition to those mentioned in the footnotes, the following works are also helpful with specific aspects of survey coding.

Quantitative coding for survey analysis is a highly developed field with many books at many levels. One of the best, though quite advanced, is the third edition (2011) of DeVellis's *Scale Development: Theory and Applications*.

Coding open-ended survey questions is most easily approached as a form of textual or content analysis. A good general text is Krippendorff's (2004) *Content Analysis: An Introduction to Its Methodology*.

A good short book that discusses several of the topics in this chapter is Blair, Czaja, and Blair's (2013) *Designing Surveys*. Chapter 9 on "reducing sources of error in data collection" is particularly helpful.

We conclude our recommendations with three fairly technical, but quite indispensable, articles on aspects of survey coding and measurement. They all focus on the most important question: How does one write *valid* questions—questions the answers to which accurately tap into respondents' beliefs? On the issue of social desirability bias—whether respondents fake answers and how to detect and correct for this if they do—see Ziegler and Buehner, "Modeling Socially Desirable Responding and Its Effects" (2009). On consulting with members of the target population, specifically by using focus group interviews, to improve the content validity of survey questions, see Vogt, King, and King, "Focus Groups in Psychological Assessment: Enhancing Content Validity by Consulting Members of the Target Population" (2004). Finally, King and Wand, in "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes" (2007), very persuasively make the case for a method—vignettes—currently being used by the World Health Organization to improve the validity and cross-cultural comparability of survey responses. This issue is particularly important when respondents interpret identical questions in different ways. King and colleagues explain methods for discovering and correcting for this complication, principally through interpreting responses to vignette-based questions. We know of no single article on survey research validity more likely to repay readers' efforts.*

*See also King, Murray, Salomon, and Tandon (2004).

CHAPTER 1 SUMMARY TABLE

CONSTRUCTING AN EFFECTIVE QUESTIONNAIRE	
When linking survey questions to research questions (pp. 24–26)	• Examine your research questions to determine whether you will collect objective or subjective data—or both. • Determine your research variables and make sure you have sufficient survey questions to answer them. • Use your survey questions to operationalize your variables.
When to use questions from previous surveys (p. 26)	• When you are examining variables studied previously by other researchers. • When you want to study whether change has occurred since a previous survey study was conducted.
When choosing question formats (pp. 27–29)	• Choose between open-ended and forced-choice formats, considering research questions, variables, and resources. • For forced-choice, choose among response types (e.g., multiple choice, Likert scales). • Consider when neutral and does-not-apply options are appropriate and/or necessary.
When redrafting questions to improve quality (pp. 31–33)	• Use a checklist. • Conduct focus group research. • Have a panel of experts review your draft. • Revisit your research questions and analysis plan. • Interview potential respondents about survey quality. • Pilot-test the survey with a sample of real respondents. • Conduct survey experiments.

CODING RESPONDENTS' ANSWERS TO QUESTIONS	
When to make a composite scale across related questions (p. 34)	• When you want to increase reliability and validity.
When you want to know whether items measure the same aspects of a variable (pp. 35–36)	• For most surveys, use Cronbach's alpha as a correlational measure of consistency. • For surveys with numerous questions per variable, use factor analysis.
When to report measures as interval or rank order (p. 36)	• Compute both ways, reporting both ordinal and interval results.