3 DESIGNING RESEARCH

Concepts, Hypotheses, and Measurement

> It is a capital mistake to theorize before you have all the evidence. It biases the judgment.
>
> —*Sherlock Holmes (by way of Sir Arthur Conan Doyle, writer)*

LEARNING GOALS

Central to doing survey research is understanding the idea of operationalization and how to go from ideas to concepts to variables. Learning the various levels of measurement is also essential for analyzing data. This chapter shows how to write hypotheses using independent and dependent variables and how to evaluate the reliability and validity of measures. By the end of the chapter, you should be able to distinguish the different levels of measurement (nominal, ordinal, and interval/ratio), discuss the various kinds of reliability and validity, and create one-directional, two-directional, and null hypotheses.

After we select a topic and review the literature, we are ready to begin constructing a research plan. A research design serves as a blueprint for the project and must be detailed when proposing a topic for a thesis or applying for a grant. A *research design* involves several stages: (1) developing concepts that are derived from ideas, theories, or prior research; (2) taking those concepts and translating them into measurable variables (operationalizing concepts); (3) selecting the most appropriate research method to gather data (surveys, experiments, field methods, content analysis, etc.) on the basis of the goals of the project (to describe, explain, predict, explore, or evaluate); (4) choosing a sampling strategy for deciding whom or what we want to study (*the units of analysis*) and over what period of time (longitudinal across time or a one-time cross-sectional study); (5) planning how to collect the data and who will

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

do it; (6) deciding on the relevant statistical and analytical tools to make sense of the findings and observations; and (7) describing plans for interpreting and analyzing the results and writing a final report, article, or policy recommendation. A detailed budget should also be included as part of a research design (especially for a grant or funding agency) which specifies everything from the costs of duplicating questionnaires to phone calls, supplies, salaries for researchers and those doing the data collection, computer data entry and software, travel expenses, online hosting fees, and other related items. The remaining chapters of the book explore these steps in the research design process as they apply in particular to survey research methods.

VARIABLES AND HYPOTHESES

Creating a research design or systematic plan is essential for carrying out a scientific study. By carefully specifying the steps necessary for researching a topic, we avoid many of the pitfalls of everyday thinking described in Chapter 1. The first important phase in this process is formulating ways to measure ideas and concepts with accuracy and consistency.

Concepts

Research is guided by a set of questions composed of concepts connected to your topic. These concepts may have been uncovered in a review of previous research and theories or developed from your own intellectual curiosity and knowledge. A *concept* is an idea, a general mental formulation summarizing specific occurrences, such as "gender" representing masculinity and femininity, or "age" summarizing specific instances of the idea of time (youth, middle age, elderly). A concept can be defined with a dictionary to produce a common usage of the word, and a search of scientific publications can result in a more suitable definition for the appropriate field of study. Otherwise, the meanings of concepts can be quite subjective and made more difficult to measure, especially for more abstract ones. Some concepts are specific and concrete, such as "height" or "academic major," while others, sometimes called *constructs*, are more complex, abstract, or difficult to define, such as "happiness" or "anomie." Ask many people to define the concept of "love," and you'll get everything from operas to paintings to poetry to puppies.

Conceptualization must occur for research to begin, and what we mean by the ideas and terms used in our study should be explicitly stated. For example, if you are interested in studying gender differences in phobias, you need to be clear about your conceptualizations. Psychologists define the concept of "phobia" as a persistent illogical fear, while sociologists define "gender" as the prescribed roles that men and women are to follow in a particular culture. A good way to start a project, then, is to list the concepts you are interested in studying and then find relevant popular and scientific definitions for them.

Variables and Values

The next step is to take the concepts of the research topic and translate them into something measurable. When this is done, these concepts are called *variables* to signify the variation that might exist in the concept. Concepts that have only one fixed meaning, such as the concept of pi (π) in mathematics, are called *constants* since they don't vary. Most concepts have multiple categories or *values* to represent the variability of the concept. "Political views" typically has several values, such as far left, liberal, middle of the road, conservative, and far right, and "religion" can vary across dozens of categories. The subjective concept (or construct) of happiness, for example, can be transformed into a variable with values ranging from very unhappy to very happy, and height could range from very short to very tall, or from 2 feet tall to 7 feet tall, depending on how you decide to measure it. Occasionally in a study, what normally is a variable can become a constant, for example, when only women complete the survey. In this case, sex becomes a constant for this particular study and can no longer be used as a variable for further data analysis.

Once concepts have been defined and translated into variables, they require some specification about how they are to be measured, or what is called the process of *operationalization*, as described later in this chapter. Measurable variables form the basis of questionnaire items that guide the collection of data. Questionnaire items are the most specific form of an operationalization and represent what the researchers believe are good indicators of the concept. Chapter 4 focuses on writing questionnaires.

Hypotheses

Central to a research design is the construction of research questions and hypotheses to guide the project. Untested statements that specify a relationship between two or more variables are called *hypotheses*. A hypothesis is a hunch derived from an informed reading of the literature, a theory, or personal observations and experience, and it must be capable of being tested. For example, a study might hypothesize that there is a relationship between employees' job satisfaction scores and the quality of the benefits provided, such as health care, vacation time, and childcare facilities. Each of these concepts is a measurable variable, and a hypothesis spells out the potential association relating them to one another (see Box 3.1).

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

BOX 3.1 CONCEPTS, HYPOTHESES, AND VARIABLES

Consider the following example from a published academic article. Note how concepts are introduced and defined, how they become variables in hypotheses, how they are then operationalized, and which ones are independent and dependent.

Vendemia et al. (2017) wondered why many college students "friend" people on Facebook they dislike or find annoying. Previous research suggested that people solicit hundreds of others on social networking sites (SNS) they are superficially acquainted with compared with the ten to 20 people they might consider close relationships in face-to-face social networks offline. But why do people connect with so many more people online whom they don't really like?

On the basis of literature reviewed and theory, the researchers decided to explore "relational anxiety" as a possible explanation. Relational anxiety is a *concept* in the psychological literature on attachment styles which begin in infant-caregiver relationships. Vendemia et al. (2017: 31) say it is *defined* in previous studies as anxiety and avoidance, and they will focus primarily on "the anxiety dimension, which assesses the uncertainty individuals feel in close relationships." Perhaps this aspect might explain why people monitor disliked and annoying "friends" on Facebook.

As a result, Vendemia et al. (2017: 31) hypothesized: "Relational anxiety is positively associated with being 'friends' with people you do not like on Facebook" and "Relational anxiety is positively associated with reading the posts of people you find annoying on Facebook." The independent variable is relational anxiety, and friending disliked people and reading posts of annoying people are the "Facebook behavior" dependent variables.

The next step is to operationalize the independent and dependent variables. For "relational anxiety," the researchers decided to use items from an "anxious attachment scale" developed by others, such as "I worry about being abandoned." Facebook behavior was measured with yes/no questions: "Are you friends with people on Facebook, even though you do not like them?" and "Are there people on Facebook whose postings you actively read even though you find their postings annoying?" Among other findings, the results indicated that the relational anxiety independent variable was a significant predictor of choosing Facebook friends who they disliked and actively reading postings of annoying others, the dependent variables. The more anxiety about relationships these college students had, the more likely they would "friend" and monitor people they really didn't care for!

How would you operationalize these concepts if you were to do a similar study?

Independent and Dependent Variables

One goal of research is to explain why a particular variable varies; why isn't it a constant? Why don't people vote for the same candidate? What is the outcome of a new benefit package in the workplace? Why do people vary in height or hold different

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

attitudes toward equal rights for gays and lesbians? The outcomes we are seeking to understand are called the *dependent variables*, and we hypothesize that their variability in our sample of respondents *depends on* particular explanations or causes. The explanations or causes (or predictors, if you are trying to predict variations in the dependent variables) are called the *independent variables*.

There is nothing inherent in a variable that makes it dependent or independent; it often is a function of the research question or hypothesis we develop. What is a dependent variable in one hypothesis could very well become an independent variable in another hypothesis or study. For example, you might be investigating whether the variation in the number of hours studying (independent variable) in a survey of college students affects differences in grade point averages (dependent variable), but later on you are interested in the reason why study hours differ so much (dependent variable) and whether this variation in your sample is due to involvement in work and extracurricular activities (independent variables).

Positive, Inverse, One-Directional, Two-Directional, and Null Hypotheses

Hypotheses should be clearly listed before you proceed with a study. However, it is not necessary to have hypotheses before conducting research; sometimes we create simple statements in question form to guide the research. These research questions can focus on describing information, such as “How many respondents in my study are over the age of 65 and under 25?” or “What is the average score on the national reading test for the local elementary school?” Other questions attempt to explain or predict relationships or differences among concepts, such as “Are variations in levels of self-esteem related to height and weight?”

A project on grades and study habits might frame the research goals in any number of ways. Consider these different versions of stating research questions and hypotheses:

1. I wonder if the number of hours studying per week is related to grades at the end of the semester.
2. If the number of hours studying per week is low, then grades are low.
3. The higher the number of hours studying per week, the higher the grades.
4. The relationship between the number of hours studying per week and grades at the end of the semester differs between men and women.
5. There is no relationship between the number of hours studying per week and grades at the end of the semester.

The first statement is simply a research question providing a hunch about a possible relationship between the two variables. The second statement is an "if . . . then" format that specifies a particular direction to a hypothesized relationship. In this case, it is called a *positive one-directional* (or *one-tailed*) hypothesis because, as the independent variable decreases in a particular way (study hours go lower), then the dependent also decreases in the same direction (grades go lower). The statement is considered *positive* because both variables are hypothesized to change together, or co-vary, in the same direction. Conversely, the third statement puts it a different way: Those who study more get higher grades. This is also a positive one-directional hypothesis.

Note that a relationship is not hypothesized about one particular person but about grouped (or aggregated) people. Hypotheses should not be written as "If a person studies less, his or her grades are lower" since the research is not designed to get information about any one individual person. The hypothesis is proposing that across the hundreds of people sampled, those who study less tend to have lower grades. Any conclusions or predictions about one person cannot be made.

A *negative* or *inverse* one-directional hypothesis states that studying for more hours is related to lower grades—a scary thought indeed! It goes in the opposite or inverse direction: As one variable increases across the sample of people in the study, the other variable decreases. If we do not want to specify ahead of time any direction (positive or inverse), then we write *two-directional* (or *two-tailed*) *hypotheses*, such as "there is a relationship between hours studying and grades." Note that the exact relationship remains unstated, and the possibility that it could be a positive one (more hours studying relates to higher grades; fewer hours studying relates to lower grades) or a negative one (more hours studying leads to lower grades; fewer hours studying leads to higher grades) is left open. We have two opportunities to support or refute the hypothesis.

The fourth hypothesis is an example in which three variables—hours studying, grades, and sex—are involved. The expectation is that the amount of time studying and the grades obtained will not be the same for men and women. It is not stated ahead of time whether the relationship will apply only to men and not to women, or the other way around. Thus, it is a two-tailed hypothesis. Researchers introduce a third variable in a hypothesis when they are interested in multivariate analysis to see if a relationship between two variables holds up under the conditions of a third or control variable or if they want to assess the impact of two or more independent variables together affecting the outcome variable. Chapter 9 discusses the process of elaborating relationships with multiple variables.

The fifth example illustrates another way of writing a two-directional hypothesis: the *null hypothesis*, which assumes that *no* relationship exists between the variables. It is a common form of stating hypotheses when using inferential statistics, since the

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

logic of research is that you do not directly prove a hypothesis: you either reject or fail to reject (that is, you accept) a null hypothesis. Here's an explanation in terms of logic that may sound like something out of *Alice in Wonderland*. When you accept no relationship, you essentially have failed to prove that the hypothesis is incorrect; you have failed to reject it. You have not demonstrated a statistically significant relationship between the variables. When you reject the null hypothesis, you accept the alternative that there is a relationship between the variables. You reject the idea that there is no relationship. Note that you haven't directly proved a relationship; you have instead disproved that there is no relationship! (see Box 3.2).

Rejecting the null hypothesis is a statistically significant event and worthy of publication, Twitter announcements, and perhaps further funding. It is not necessarily a good or bad thing; no value judgment is implied when rejecting or accepting a hypothesis—you may have found a significant relationship between hours watching television per week and an increase in grades, something you were hoping would not happen—but it is something that gathers attention. The goals of your research guide you in interpreting the results, and the hypotheses assist you in scientifically designing your study.

BOX 3.2
THE LOGIC OF THE NULL HYPOTHESIS

In a sense, a null hypothesis exclaims, "Show me that I am wrong; I say there is no relationship," and you find that either you are wrong ("go ahead and reject my hypothesis; I guess there is a relationship after all") or you have withstood the challenge ("accept the hypothesis of no difference; I told you so!"). But in neither case have you proved the hypothesized relationship directly. Rather, the null hypothesis has failed to be disproved (it is accepted as is, as no relationship), or you have disproved it (it is rejected and a relationship exists; hold a press conference!).

Here's an example in logic: You state, "There is no such thing as a chicken with purple polka dots." You can do a census of all chickens in the world and see if this is so; if there are no purple polka-dotted chickens in the entire population, then you have indeed proved your point. Realistically, you aren't going to check every chicken in the world, so you take instead a sample of chickens (more on sampling in Chapter 5), and lo and behold, none of the chickens in your sample have purple polka dots. Then maybe you find another sample, and once again there are no purple polka-dotted chickens.

You have failed to disprove or reject your statement: You accept it and infer that there are no purple polka-dotted chickens in the world on the basis of your samples. All you can scientifically and logically conclude is that you accepted your null statement that no purple polka-dotted chickens exist. However, you did not prove directly that there exists no colorful poultry like this. Somewhere in some hidden chicken farm, there might be a purple-dotted chicken pecking away. If you found one, you would reject the null, and state the *alternative hypothesis* that there are some. Call in the press for a major news story and marketing plan to get your photo of the poor chicken as a Snapchat filter!

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Defining and Operationalizing Variables

With a set of questions or hypotheses, list the variables in your study and specify their conceptual and nominal definitions. Be attentive to which ones are independent variables and which are dependent. It is important that the concepts underlying the variables are clearly defined before you go to the next step in figuring out how to measure them. Consider a study on political attitudes and ethnicity/race and the following hypothesis: There is a difference among various racial/ethnic groups about their opinions of the last presidential election or, as stated in the null form, there is no difference in opinions about the election among various racial/ethnic groups. You assume that people's views on the outcome of that contest depend on their race/ethnicity. You list the two variables: racial/ethnic group (the independent variable) and opinions about the election (the dependent variable). This is a two-directional hypothesis since you are not predicting ahead of time which category of the race/ethnicity variable will relate to which set of attitudes about the election.

Now you must define these concepts: What does race/ethnicity represent, and which attitudes do you want to assess? Do you really mean their attitudes, or were you thinking to include a behavioral dimension, such as for whom they voted? Once you decide on the concepts you intend to study and turn them into variables by concretely defining them in measurable ways, the next step is to specify exactly how you are going to measure these variables. The objective is to develop appropriate indicators for the variables and concepts you are studying. A concept can be measured using any number of ways to show its variability. The concept of "sadness" can be indicated by using one "Are you sad?" question in a survey, or it can be assessed through multiple indicators, such as a set of ten items that form a "sadness" scale. A *scale* is a composite measure combining responses to many questions believed to assess the same concept or variable.

This process of specifying the indicators of a concept is called operationalization, because we are describing the operations or procedures it will take to assign values to the variables. Think about the time someone told you that a movie was really interesting, and you asked for clarification on what "really interesting" meant. What you were doing was asking for a definition, asking how your friend measured the quality of the movie, and asking what operations he or she used to arrive at the value of "really interesting" for this attitude variable. The values for the variable could have ranged from "really bad" to "really excellent," with "really interesting" somewhere in the middle. Or "really interesting" could have been the highest accolade in your friend's measurement scale. The way a concept is operationalized needs to be spelled out in detail in order to be scientific and not lapse into the errors of everyday thinking as described in Chapter 1.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Similarly, if you define the concept of race/ethnicity as a category with which a respondent identifies, then you might operationalize the variable as a person self-disclosing as African American, Asian/Pacific Islander, Hispanic, white, Native American, mixed race/ethnicity, or other. Or it could be measured as the category that was listed on a birth certificate. The actual item we develop for a questionnaire is composed of the categories of the variable race/ethnicity.

Let's say "attitude toward a political election" is defined as how intensely respondents feel about the election results and whether they perceive them as fair or not fair. Attitude will be operationalized by developing a 5-point measure from "strongly agree" to "strongly disagree" to assess what the respondents believe about such statements as "The outcome of the mayoral election was achieved fairly" and "The campaign used too many negative advertisements."

The conceptual definitions and operationalizations are important to describe in a study because if others wish to replicate your research, it is essential that they know the steps you took to measure your variables. When they review the literature and find your study, they might agree or disagree with the way you defined and measured the concepts. So long as you are clear with what you did, you have provided readers with enough information to make a fair assessment of your work (see Box 3.3).

BOX 3.3
CONCEPTUALIZING AND OPERATIONALIZING "UNEMPLOYMENT RATE"

Politicians use a variety of economic and social indicators to suggest how well their policies are working out. One of the most controversial figures is joblessness or unemployment rates. Think for a moment how you would conceptualize this indicator and how you would go about measuring it. In the U.S. on the first Friday of every month, the Labor Department releases a jobs report. The department's *standard definition of unemployment* "counts only people who say they want a job and have looked for one in the last month—meaning that millions of Americans who dropped out of the labor force . . . do not count as unemployed" (Irwin 2017: B4). Yet there are dozens of pages of figures in the monthly report that can be used to calculate unemployment differently.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

BOX 3.3 CONTINUED

Consider these alternative ways to operationalize and conceptualize unemployment, in addition to the "standard definition" (Irwin 2017; Bureau of Labor Statistics 2015):

1. People who want a job but have not looked for one in the last month
2. Part-time workers wanting full-time work
3. The over-16-year-old population who is not working
4. The 25- to 54-year-old population who are not working, to eliminate students and retired persons
5. People collecting unemployment insurance
6. People working or looking for work who are not institutionalized (mental hospitals, residential nursing facilities) or in the armed forces
7. People who don't have a job and are not looking for one.

Not all concepts are ready to measure without some discussion about the intention and goals of the project, sometimes guided by political concerns. All of these methods to operationalize and define unemployment have been calculated. Clearly articulating the definitions of the concepts and how to measure them must be systematically detailed in any research proposal and presentation. How would you define unemployment?

The more abstract the concept, the harder it is to develop an operationalization. How would we measure the constructs of "love" and "alienation," for example? On the other hand, there are only so many ways we can operationalize "income" in a study: Do we ask for income before or after taxes, with or without investments, per month or per year, for an individual or the entire household? But differences in the way we measure concepts can affect the outcomes and interpretation of the findings.

LEVELS OF MEASUREMENT

Operationalizing variables involves specifying *levels of measurement* and considers how reliable and valid the measures are. For constructing a questionnaire, it is important to understand the many ways variables can be written. Not only can differences in measurement affect the statistics used when analyzing the data (as discussed in Chapter 6), but they also determine the various kinds of information you actually get. Researchers make four general distinctions in measuring variables: nominal, ordinal, interval, and ratio measurements. In addition, variables are discrete or continuous (see examples in Box 3.4).

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

BOX 3.4
LEVELS OF MEASUREMENT

Consider these excerpts from published studies and see how researchers decided to operationalize the variables with different levels of measurement: nominal, ordinal, and interval/ratio.

1. “Sexual identity, measures respondents’ self-identified orientation, and is asked as, ‘Which of the following best describes you?’—heterosexual, gay/lesbian, or bisexual. . . . Race of the respondent is measured as white, black, or other race. . . . Region of the United States where the respondent lives includes the four major Census regions: Northeast, Midwest, South, and West.” (Mize 2016: 1140–1141)
2. “Our main independent variables are educational attainment and race. We code educational attainment into four categories: high school or less, some college (including associate’s degree), college graduate (bachelor’s degree), and advanced degree. We include four racial groups in our sample, which we code as mutually exclusive—white, black, Hispanic, and Asian (which includes Pacific Islanders), with whites as the reference group.” (Yavorsky et al. 2016: 741)
3. “How much do you like the following types of music: chart music, hip hop, R&B, rock, heavy metal, punk/alternative, dance/house, techno, hard-house, and classic music. Participants rated their responses on a five-point Likert scale with response categories 1 (dislike very much), 2 (dislike), 3 (do not dislike or like), 4 (like), and 5 (like very much).” (ter Bogt et al. 2010: 851)
4. “School poverty is measured by the proportion of students who receive free or reduce priced lunches. School size is measured by total student enrollment.” (Peguero 2016: 9)

For the first example, each of these is a nominal measurement: Sexual identity is represented by three categories in no particular order, race with three categories, and region of the country with four. If these categories were coded with numerals, such as 1 = Northeast, 2 = Midwest, 3 = South, 4 = West, it would not mean that West was twice as much as Midwest, or that Northeast is less than South. These are not coded with actual numbers that have any mathematical properties.

The second example includes the variables “race” and “educational attainment.” Categories for education are in order but do not represent equal intervals, so it’s an ordinal measurement. The other item about race is often called a dummy or dichotomous variable with two mutually exclusive nominal categories each assigned an arbitrary numeral: 1 is for those respondents who are the reference group, in this case white, and 0 is for each of the other racial categories such as 0 = Asian, 1 = white; 0 = black, 1 = white, and so forth.

The third example is technically a *discrete ordinal* measure in which the numerical values are assigned in order from “dislike very much” to “like very much” with three other answers in between. It is arbitrary that “dislike very much” is assigned a 1; it could have been a 4. However, once the numbering starts, the remaining categories must be assigned a numeral in order to represent the increase or decrease in opinions.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

BOX 3.4 CONTINUED

Ordinal values can be described as "more" or "less" and "larger" or "smaller" than those below and above. These comparative properties are central to the idea of ordinal measures. If the categories are viewed as equal-appearing intervals—that is, if the gap from "dislike very much" to "dislike" is seen as the same as that between "like" and "like very much"—then this ordinal measure could be treated as interval/ratio for statistical purposes (as discussed in Chapters 6 through 9). These are often referred to as Likert scales.

The last example uses two interval/ratio measures as indicated by actual numbers: the proportion of students' reduce price lunches (derived by using mathematical calculations on a count of the number of these lunches) and enrollment figures.

Discrete and Continuous Measures

A *discrete* variable is measured with values that do not contain additional information between those values, such as the number of children in a family. There are only exact numbers of children; there is nothing in between the counting units. Do you know someone with 1.3 children, or 3.75 kids? Type of car owned can be a Honda, a Ford, a Tesla, and so on, but there is no category between a Honda and a Ford. A *continuous* variable, on the other hand, uses values that have information between those values. Time is measured continuously: The counting units of 1:15 P.M. and 1:16 P.M. contain seconds between them. If your watch has a second hand, you can measure them. In track races, you break down the units between minutes and seconds even further to hundredths of a second, and computers can calculate time in picoseconds. We arbitrarily say, "It's around 1:15" unless you're one of those smart alecks who, when asked the time, responds, "It's 1:16 and 23 seconds" or "We've been dating for 2 months, 3 days, 14 hours, and 6 minutes!"

The distinction between discrete and continuous is useful in interpreting data since it occasionally is reported in the press that, for example, the average family size is 2.58 people. A statistic appropriate for continuous variables (a mean) was calculated for discrete data. A more important distinction, however, is the level of measurement used to operationalize the variable.

Nominal Measures

The simplest, and perhaps least useful statistically, is the nominal level of measurement. *Nominal* variables use discrete measures whose values represent named categories of classification. A measure used to list academic disciplines, such as sociology, psychology, anthropology, history, and economics, is a nominal one. A value or

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

number can be assigned to these categories (called *coding*) as a shortcut summary of the category, but it is not mathematical and no order to the categories is intended. On a questionnaire, for example, the variable "religion" can have the value 1 for "Catholic," 2 for "Protestant," 3 for "Jewish," and 4 for "Muslim," or it could be listed as 1 for "Jewish," 2 for "Catholic," and so on. These are arbitrary numerals that do not represent any type of mathematical quantity. They're like postal codes or telephone numbers. We can't add or multiply them or claim that our phone number is larger and therefore more important than someone else's. If a friend told you the type of car she drives by giving you the category number without telling you the coding scheme, it would make no sense, because it does not represent an actual value or quantity. ("I drive a 3." Huh?)

In fact, we do not need to assign a number to the categories because many computer programs allow for data analysis of text, but it is a convenient way of entering data into a computer program for later analysis. There are fewer keystrokes when typing a "4" than when typing "cultural anthropologist." When there are two categories, the variable is called a dichotomy or a *dichotomous variable*. In some statistical calculations, these two-category variables are called *dummy* variables; for example, the variable "work status" can have one category value "employed" and another value "not employed."

Ordinal Measures

When the category values for a variable are in sequence, the measurement is considered *ordinal*. This is also a discrete measure but one in which the values increase or decrease in a particular order. In this case, it does make a difference which one comes before or after the other. For example, we might ask respondents to indicate their shirt size (the variable) in terms of small, medium, or large (the values). They go in order; we cannot say medium, small, large, and then assign the numbers 1, 2, and 3 to those words since they are not in sequence. If something is in category 1, it must be less than 2, which in turn must be less than 3. Or we could assign numbers the other way around: 1 can be the most, 2 in the middle, and 3 the least—just so long as we keep going in rank order once we start the numbering.

When we are asked to line up in size places, we are using an ordinal measure to determine who is shortest and who is taller and who is tallest. We don't measure with a ruler to get the exact inches; it's just a visual estimate of degree, usually greater than or lesser than. When you ask respondents to disclose their yearly income using such discrete categories as "under $30,000," "from $30,000 to $35,000," "from $35,001 to $40,000," and "over $40,000," you are creating an ordinal variable. These categories are in sequence and can be coded with ordered numerals showing increasing degrees of wealth, such as 1 for “under $30,000,” 2 for “from $30,000 to $35,000,” and so on.

Yet the ordered number still doesn’t tell you the answer without first consulting a coding sheet of information. If someone responded to a question about how much he earned last year by saying, “I earned 3,” you couldn’t tell what category of income that is just simply by the numeric value. All you know is that he is earning more (or less, depending on the order) than those in the first two categories. Don’t get coding numerals confused with the numerical values of the responses.

Interval and Ratio Measures

When the value tells you everything you need to know about the variable, you are operationalizing it with an *interval* level of measurement. These are measurements whose numbers are in order with equal size intervals and that have no absolute or fixed zero as a starting point. Temperature, for example, does not begin at zero; in fact, zero degrees is not the absence of temperature, it is an actual temperature. Yet the number itself tells you what you need to know. If you ask how cold or hot it is outside, and the response is 25, you do not need to look up what category that numeric value represents. It is an actual number that can be treated mathematically (that is, added and subtracted). It is 10 degrees more (remember, it has the properties of ordinal measures) than 15 degrees, and 5 degrees less than 30. However, you cannot say that 20 degrees is twice as warm as 10, or half as warm as 40, since there is no absolute zero point to calculate a ratio.

When there is an absolute zero, it is called a *ratio* measure. Now a zero value represents the absence of something and the start of the numbering system. There are no negative numbers below zero on this type of measure. Thus, zero weight is the absence of all weight, if that were possible; you certainly cannot have a negative weight. Ratio measures allow you to use the other mathematical operations (multiplication and division) to achieve ratios and conclude that someone who weighs 200 pounds is twice as heavy as someone who weighs 100 pounds, or that $10 an hour is half as much as $20 an hour. Note that this is different from a measure that is used to assess the severity of an earthquake. The Richter scale is not a ratio measure, but one in which the units are based on logarithmic calculations, so that an earthquake of 7 compared with one of 6 is an increase in amplitude by 10. So you can’t always assume that a measure using actual numbers is a ratio or interval one.

Ratio measures are so similar to interval measures, except for the presence or absence of zero, that throughout the book, such variables are called *interval/ratio*. They are often continuous measures, although discrete interval/ratio measures also exist, such as number of books in the library, children in households, houses on a

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

block, or any other similar counting situations. The most advanced and important statistics require interval/ratio levels of measurement.

Other Measurement Considerations

Not every measurement falls neatly into these four distinct classifications. Dichotomous variables technically have discrete nominal measures. However, an order exists when there are only two values: Assigning an arbitrary 1 to "home owners" and a 2 to "renters" suggests that renters are "higher" than home owners for the duration of the data analysis in this particular study. Therefore, dichotomies can sometimes be treated statistically as ordinal or interval/ratio measures. Another example is an intensity measure, such as a 5-point range (often called a *Likert scale*) where 1 is "strongly agree" and 5 is "strongly disagree." These are ordinal measures, but most researchers treat intensity scales as interval/ratio measures when the amount of agreement or disagreement is assumed to vary in equal intervals along the points of the measure.

It is ideal to strive for the highest level of measurement when constructing items for a questionnaire. The levels of measurement are themselves ordinal, with nominal the least powerful, ordinal ones in the middle, and interval/ratio the most complex. Interval/ratio measures have all the properties of the ones below them. We can always recode interval/ratio data into an ordinal measure, but not the other way around. For example, you can ask respondents their ages in terms of years and months (a continuous interval/ratio measure) and then later move them into "young," "middle age," and "elderly" ordinal categories or group them in ordinal age ranges such as "under 20," "from 20 to 30," "from 31 to 40," and "over 40." However, if you operationalize age in either of these last two discrete ordinal ways, you can never get an exact list of ages or calculate a mathematical average age. If 100 respondents select "over 40" on your questionnaire, you'll never know if all 100 are each 44 years of age, for example, or if they represent a range of ages from 41 to 99.

SCALES AND INDEXES

To measure a complex concept, researchers often construct scales and indexes (or indices). These words are often used interchangeably, but some distinctions can be made. An *index* is a set of items that measure some underlying and shared concept. Think about the Dow Jones Industrial Average, which is a set of numerous stocks combined to create a single number. Now apply that idea to a concept such as "happiness" or "prejudice" or "self-esteem." Creating an index is developing a set of items that together serve as indicators of the underlying concept you are trying to measure.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Inter-item correlations are mathematically calculated to determine how well the individual items in the set relate to each other and to the overall concept being measured. For example, imagine a "wellness" index created to measure respondents' attitudes toward physical and mental well-being. A set of questions might include such items as "Most days I feel happy about my physical appearance," "I believe exercise is important for my mental health," and "I often feel lonely."

Perhaps a dozen items like these are written, and the respondents are then asked to indicate how well each question represents their feelings on a scale of 1 to 7, where 1 is "not at all" and 7 is "very well." Researchers would calculate an overall total health index score (an interval/ratio measure) by summing together the responses on each item. In this example, with 12 items, the scores could range from a low of 12 to a high of 84. During the index development stage when the questions are pretested, a researcher would calculate how well each item correlates with the others and with the overall score on the health index. Those items that work well together in measuring attitudes toward health are kept, others are dropped, and perhaps newer ones are added. As discussed in the next section, various ways of determining validity and reliability are necessary when constructing items, scales, and indexes.

A *scale* is a set of items that are ordered in some sequence and that have been designed to measure a unidimensional or multidimensional concept. Usually, a pattern is sought from the responses to a set of items, rather than a simple summation of the individual item scores, as with indexes. Take, for example, Guttman scales, in which agreement with a particular item indicates agreement with all the items that come earlier in the ordered set. Respondents would be asked a series of items indicating (1) if they would be comfortable if someone of a different race/ethnicity lived in their neighborhood, then (2) if they would be comfortable with having someone of a different race/ethnicity as a next-door neighbor, and then (3) if they would be comfortable with having their adult child marry someone of a different race/ethnicity, and so on. The items are ordered so that agreement with the third one would also indicate a strong likelihood of agreement with the first two statements.

However, people usually refer to a single item that is measured ordinally as a scale, such as questions rated on a "scale" of 1 to 10 or assessed with a Likert "scale." A Likert scale reflects a level of preference or opinion, typically measured on a 5-point ordinal scale such as "strongly agree," "somewhat agree," "neither agree nor disagree," "somewhat disagree," and "strongly disagree." Technically, these individual items are not a scale in the sense of the outcome of the complex process of "scaling." These Likert items or rating scales are often combined to form an index, although such combinations of measures are sometimes called "summated scales." By now, you may be getting the correct impression that the words "index" and "scale" are used interchangeably by many people!

Today, researchers tend to construct mostly indexes (or summated scales) and only occasionally develop cumulative scales like the Guttman scale. The word "scale" is more often used than *index*, and the original differences between them have become blurred over time. Whatever it is called, the key point is to remember that the measurement of complex ideas and concepts ideally requires more than a single item. Combining questions into an overall measurement ("scaling") usually requires advance planning, pretesting, and the statistical establishment of reliability and validity. However, it is more common and much easier to create an index by summing a set of items that have been developed to measure a particular concept even after the data have been collected and statistically analyzed. When you learn to write questions in Chapter 4, consider creating a series of items that might work together in measuring a particular concept. The resulting total score is an interval/ratio measure that can be used in many advanced statistics.

ACCURACY AND CONSISTENCY IN MEASUREMENT

We can construct the best measures in the world, but if they aren't accurate and consistent, our findings cannot be trusted. One of the major sources of error in studies is poor quality of the measurements. Operationalizing variables requires attention to two core concepts of research methodology, namely, validity and reliability. Often these cannot be determined until we have already completed data collection, unless we do pilot studies that test our items first before developing a final version or use (with permission) items previously written for other studies that already have demonstrated validity and reliability (see examples in Box 3.5).

BOX 3.5
LOOKING AT RELIABILITY AND VALIDITY

Here are some examples from published research that discuss issues of reliability and validity in the measurements of the variables used in the various studies:

1. "A key advantage of the NAB [Neuropsychological Assessment Battery] Naming Test is the availability of two alternate forms, which allows for serial administration to minimize the likelihood of practice effects.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

BOX 3.5 CONTINUED

Psychometric data from the test publisher indicate that the NAB Naming Test has adequate internal consistency (Form 1: $\alpha = .79$; Form 2: $\alpha = .73$), stability over an average test-retest period of 6 months ($r = .70$), and alternate-forms reliability (generalizability coefficient = .72).” (Sachs et al. 2016: 630)

2. “Concurrent validation procedures were employed, using a sample of African American precollege students, to determine the extent to which scale scores obtained from the first edition of the Learning and Study Strategies Inventory (LASSI) were appropriate for diagnostic purposes.” (Flowers et al. 2011: 1)
3. The Columbia Mental Maturity Scale, published by Harcourt Brace Jovanovich, estimates the general reasoning ability of children from 3.5 to 10 years of age and assists educators in selecting appropriate curriculum materials and learning tasks: “Split-half reliability across all levels approaches .90 and test-retest reliability is approximately .85. Concurrent validity with the Stanford Achievement Test, the Otis-Lennon Mental Ability Test, and the Stanford-Binet Intelligence Test ranges from the .30s to the .60s.” (Retrieved from http://ericae.net/eac/eac0069.htm)
4. The 40-item Friendship Quality Questionnaire—Revised “had high internal consistency at both Time 1 ($\alpha = .96$) and Time 2 ($\alpha = .95$) and test-retest reliability of .38 ($p < .01$).” (Kingery et al. 2011: 224)

The first example describes a very comprehensive discussion of reliability for a published measurement tool used by neuropsychologists. Alpha statistics (where perfect reliability is 1.0 correlation and no reliability is 0.0; see Chapter 7) indicate that the NAB test is fairly reliable in terms of internal consistency. *A test-retest consistency* procedure establishes the reliability of the NAB over a six-month period and the reliability of parallel forms of the assessment tool is established.

The second example focuses on determining whether a published questionnaire used to measure study skills is accurate in measuring actual academic achievement. This is *concurrent* or *criterion validity* since another established measure (in this case, scores on the ACT college entrance test, according to the article) is used as a criterion to assess the accuracy of the LASSI study skill inventory.

The third item presents reliability and validity information about a commercial standardized test available for schools to purchase for developmental assessment purposes. Note that two types of reliability were measured: *split-half* and *test-retest*. Both demonstrate high reliability (correlations approaching 1.0 are considered high). By comparing scores on this standardized test with scores on three other achievement, mental ability, and intelligence tests, researchers were able to demonstrate concurrent validity, although the correlations were moderate.

The last excerpt from a published academic article illustrates how the 40 items that make up this questionnaire on friendship quality showed strong reliability (consistency) among themselves but had a somewhat low reliability when repeated over time.

Validity

We all need to be validated, as we say in “everyday speak”; that is, we all need to be recognized and taken for who we really are. So, too, with measurements. They need to be taken for what they really are: Are the operationalizations measuring what they are intended to measure? *Validity* is about *accuracy* and whether the operationalization is correctly indicating what it's supposed to. A ruler would be a valid measure to assess height, but a scale used to weigh yourself would not be a valid measure to assess height. Validity depends in part on what is being studied; the scale becomes an accurate tool when assessing weight in another study. There are several ways of determining if the measures you use are valid: face, content, construct, and criterion validity.

Face and Content Validity. A legitimate, but not very mathematical, way of assessing validity is to see if the measure seems to be getting the desired result. Usually a consensus develops among researchers as to whether a measure is doing what it's supposed to be doing. Take it on "*face* value" and ask, for example, does the questionnaire item about zodiac sign seem like an accurate way of indicating someone's height? It's not likely to be an accurate measure of that variable, so just on the face of it, it's not valid for assessing height. On the other hand, a question such as "how tall are you?" on its face appears to be a valid way of measuring height. Of course, the respondent might not answer truthfully, so face validity is not a perfect way of determining the accuracy of an item.

Content validity is an equally subjective way to understand how well a set of items is measuring the complexity of a concept or variable we are studying. Does the content of the items cover all the dimensions of the idea? For example, does the content of the driving test required to get a license include the range of things necessary to be a safe driver? A driving test that asked questions only about hand signals or did not require a parallel parking demonstration would not be a very accurate measure of driving skills. Its validity would be called into question. Again, consensus among researchers evaluating the measures is used to determine content validity.

Construct Validity. A better way of assessing the accuracy of a measure is to determine its *construct validity*. A construct is an abstract, complex characteristic or idea that typically has numerous ways to measure it. Take, for example, an idea of measuring student satisfaction with the university. This is not as concrete a concept as asking respondents their height.

Imagine we develop several ways of assessing satisfaction. On the basis of a statistical connection among these various measures (such as finding out that those who are dissatisfied with the quality of the teaching are also dissatisfied with the classrooms, sport and other recreation facilities, and the relationship with the local community surrounding the campus), we conclude that the items developed are measuring the abstract concept of satisfaction with some degree of accuracy. And if we found out

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

that disliking the food in the cafeteria was strong among both those who were satisfied with college and those who were dissatisfied, that item would not be as accurate an indicator of overall college satisfaction and therefore would have lower construct validity.

Construct validity is based on actual results; sometimes it is not achieved until after the data have been collected and statistically analyzed. Then this information is available for the next time someone proposes research on this topic and is operationalizing similar variables.

Criterion Validity. Another good way of determining validity, especially for constructs not easily measured, is to see if the results from an item or set of measures (a scale) are similar to some external standards or criteria. If these other criteria are available at the same time (concurrently) as the new measures are being used, then we establish *concurrent validity*.

For example, results from an item asking respondents their grade point average are compared with the official information stored by the school. Or take a more abstract variable, such as religiosity. We develop a set of questions we feel measure how religious respondents are. We give the items to a group known to be very religious already and statistically assess whether the measure indicates high religiosity for this group. If so, we have established that the items we wrote accurately clarify people's level of religiosity, and we can now use them with more confidence in our own study. The items are measuring what they were developed to measure; that is, they are accurate and valid.

Another type of criterion validity assesses how accurately our measures predict some future, rather than current, outcome. This is usually referred to as *predictive validity*. Imagine we have written a few items for a questionnaire that we feel indicate motivation to succeed in the workplace. We later track respondents to see if indeed they have been successfully promoted, received raises, and have been productive employees. We can then statistically determine if the motivation items were accurate predictors of success and, if they are, declare that they demonstrate predictive validity. We or others can now use them with some confidence and accuracy in another study on this topic or to screen potential employees, because we have established their validity.

Reliability

Just because a measure is valid doesn't necessarily mean it is reliable, and validity means little if the measure used is not reliable. Cars are a valid tool to get from one point to another around town; yet, if you've had cars like mine, they have not always

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

been the most reliable. We want to be sure that when we turn the engine on, the car starts every time. We want some stability. *Reliability* is about *consistency*; it is the expectation that there won't be different findings each time the measures are used, assuming that nothing has changed in what is being measured. For example, we would expect our weight to change little within a few minutes' time, so repeated measures (getting on and off a scale five times in a row) should indeed demonstrate a consistent value. However, if the measurement tool we use to evaluate weight yields inconsistent results, then that cheap scale you bought at the 99-cent store is not very stable or reliable. There are several ways of determining if the measures we use are reliable: test-retest, parallel form, inter-item, split-half, and inter-rater reliability.

Test-Retest Reliability. Consider a ruler made of rubber and a 25-inch-long desk. You can use the rubber ruler to measure the length of the desk; it seems to have face validity as a measuring tool because it has inches marked off along the ruler. Because the desk length does not change, repeated measures should yield the same lengths. So you try again and now the ruler tells you that the desk has grown a few inches to 28. You measure one more time and it now is 32 inches long. Either some supernatural event has just occurred or the rubber ruler has been stretched each time it has been used. Results have not been consistent each time you tested and retested with the ruler, and the reliability of the measurement tool is thus determined to be very low. Shopping tip: Don't buy rulers made out of materials that stretch!

Similar methods are used to evaluate the reliability of questionnaire items and other types of scales. If we asked respondents to evaluate their fear of crime and a few days later asked them again, only to find different results—and assuming nothing major has occurred in the media or their neighborhoods to alter those feelings in just a few days—we may be dealing with unreliable measures of fear. Many commercial and other standardized scales, such as the Rosenberg Self-Esteem Scale, the SAT, and the Stanford-Binet Intelligence Test, publish their test-retest reliability using statistical measures of correlation (discussed in Chapter 7).

Parallel Form and Inter-Item Reliability. We can also determine how consistently a concept or construct is being measured by comparing it to some equivalent measure or set of items, either externally or internally. Many commercial and standardized tests have two versions, and if the same people score approximately the same on both form A and form B, we can say there is some parallel reliability between the alternate versions. This is what you expect when a professor gives a make-up test: You hope it is the same level of difficulty as the original test!

When we compare responses to similar items within a questionnaire to see if there is consistency in the parallel measurements, it is called *inter-item reliability*.

For example, we might ask respondents how many hours they study on average in a typical week and compare this to their answers somewhere else on the questionnaire about the number of hours they study on average in a typical month and other related items. If the number of hours per month is about four times the previous answer, then we can demonstrate some reliability with these measures.

This also illustrates that the more ways we ask something, the more the potential for reliability increases. Imagine using just one question to indicate how happy people are in their romantic relationship, as opposed to asking them to respond to multiple items that are indicators of happiness. Reliability is obtained when you can demonstrate consistency among these multiple measures of happiness.

Split-Half Reliability. A popular way of statistically determining consistency is to look at internal stability by selecting a group of items developed to measure some construct or variable (a scale or index) and then comparing answers within this group. Consider a set of items you wrote to measure the political values of respondents. Let's say there are ten items. You would split the number of items in half—for example, either the first five and the last five, or the odd-numbered ones and the even-numbered ones, or randomly select two sets of five—and compare the scores of the two halves. If five of the items indicate low political involvement, then the other five should be consistent and also show low political involvement. If so, you can say the scale developed to measure political involvement is a reliable one. A statistic called Cronbach's alpha (α) is often used to assess internal consistency: The closer the correlation coefficient is to 1.0, the more reliable it is.

Inter-Rater Reliability. When researchers use open-ended items on questionnaires or gather information using interviews and other qualitative techniques, it becomes important that the data collected are interpreted in consistent ways. Content analysis and observation field research require some degree of agreement among those who are reading the data or observing. If those coding the data agree, then we can claim there is intercoder or inter-rater reliability. The interpretations of the qualitative responses are consistent among various coders or readers.

Achieving reliability and validity is part of the process of operationalizing the variables in research questions and hypotheses. And operationalization is ultimately accomplished by the process of writing a questionnaire. The reliability and validity of items and scales in a questionnaire are affected by the skill, creativity, and techniques we employ when designing everything from the survey's format to the wording of the questions. The next chapter takes you through these next steps of the research journey.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

REVIEW: WHAT DO THESE KEY TERMS MEAN?

Coding

Concept

Constants

Constructs

Continuous measures

Dependent variable

Dichotomous variable

Discrete measures

Dummy variable

Hypothesis

Independent variable

Index

Indicators

Interval measures

Levels of measurement

Likert scales

Nominal measures

Null hypothesis

One-tailed, two-tailed hypotheses

Operationalization

Ordinal measures

Positive and negative (inverse) relationships

Ratio measures

Reliability: test-retest, parallel form, inter-item, split-half, inter-rater

Research design

Scale

Units of analysis

Validity: face, content, construct, criterion, predictive, concurrent

Values

Variables

TEST YOURSELF

For each of the following hypotheses, determine which variable is independent and which is dependent, what the levels of measurement are (nominal, ordinal, or interval/ratio), and what kind of hypothesis it is (one-directional positive, one-directional inverse, or two-directional null or positive).

1. There is no relationship between education level (1 = high school graduate, 2 = some college, 3 = college graduate, 4 = graduate school) and scores on a scale measuring life satisfaction (scores range from 1 to 10, where 10 = highly satisfied with one's life).

	Which Variable?	Level of Measurement?	Type of Hypothesis?
Independent variable			
Dependent variable			

2. Men are more likely to receive higher hourly wages than women.

	Which Variable?	Level of Measurement?	Type of Hypothesis?
Independent variable			
Dependent variable			

3. There is a relationship between ethnicity/race and political party affiliation.

	Which Variable?	Level of Measurement?	Type of Hypothesis?
Independent variable			
Dependent variable			

INTERPRET: WHAT DO THESE REAL EXAMPLES TELL US?

1. What *levels of measurement* are used for each of the variables described in the following excerpt from a published study on adult children of heterosexual couples and same-sex couples? (Richards et al. 2017: 5–7)

   Parents were asked the following:

   a. "How old is the child?"
   b. "What sex is the child?" Response choices included "Male," "Female," "Other."
   c. "How important are [child name]'s religious or spiritual beliefs to him/her?" Response choices included "Not at all important," "Somewhat important," "Moderately important," "Very important," "Extremely Important."
   d. "How often are you in contact with [child name]?" Response choices included "This child lives with me," "Every day," "Once a week," "Once a month," "Several times a year," "Once a year," "Less than once a year," and "Never."
2. Here are some *hypotheses* from a published study on predictors of media multitasking in Chinese adolescents (Yang and Zhu 2016: 432). For each hypothesis, label the independent and dependent variables and describe what kind of hypothesis it is: one-directional or two-directional. How would you state these as null hypotheses?
   a. Age is correlated with media multitasking.
   b. Girls will be more likely to multitask than boys.
   c. Time management will negatively correlate with multitasking.

CONSULT: WHAT COULD BE DONE?

You are asked to serve as an advisor on a project that the local high school is starting. Administrators want to know whether students' self-esteem affects their grades and plans for college. The administrators intend to use a published

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

standardized self-esteem scale and develop new questions about grades and future plans for a survey.

1. What steps would you suggest they take? What should they do first? Begin consulting with them on a research design.
2. What would you advise them to consider when developing items for their questionnaire and when choosing a standardized scale? See if you can find published information for them about some standardized self-esteem scales and evaluate the information about their reliability and validity.
3. What are some ethical concerns that need to be considered?

DECIDE: WHAT DO YOU DO NEXT?

For your study on how diverse people develop and maintain friendships, especially on social media, respond to the following items:

1. Write five hypotheses or research questions using ten different variables. Try to write different kinds of hypotheses, such as a one-directional (one-tailed) negative (inverse) hypothesis, a positive one, or a two-directional hypothesis. Use both null and alternative hypothesis wording.
2. Make a list of all the variables in your five hypotheses, and label each one as "independent" or "dependent."
3. Say how you could operationalize each variable in your hypotheses and what level of measurement you would use.

Copyright © 2018. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.