Section 2: Ten Tools for Applying Sociology

CHAPTER 2.7: MEASUREMENT

QUICK START: In this chapter, you will learn

- What it means to measure something.

- How to measure a social variable, and at what level of measurement.

- When to use qualitative and/or quantitative measures.

- When a measurement is reliable and valid.

TERMS

Measurement	Ways to observe or indicate a person or groups' values, attitudes, beliefs and behaviors.
Indicator	A question used to measure a concept that can't be directly observed or measured.
Code	A symbol scheme, usually numbers, which provides a shorthand way of representing the values of a social variable.
Measure (noun)	Individual data collection items such as a survey question, interview question, items for observation, items to be investigated in a historical study, or items to be recorded in an experiment. Closely related to the term "variable."
Instruments or Tools	A collection of measures used in a study, such as a survey questionnaire or interview guide.
Reliability	An indication of whether a measure measures the same way every time.
Validity	An indication of whether a measure measures what it is supposed to measure.

Chapter 2.7: Measurement

Scale
A response set that identifies a numerical range for respondents to place themselves on, such as, "How much pain do you feel now, *on a scale of 1-10*?"

Pre-test
A trial run of a study's measurement instrument on a subset of the population.

Question Stem
The part of a survey measure that asks a question, such as, "In the calendar year 1999, were you victimized?"

Response Set
The part of a survey measure that provides the answers for respondents to choose from, such as, "Yes, No, or Don't Know."

Quantitative Measurement
Ways of measuring that breaks down attitudes, beliefs, values and behaviors into their component parts and marks or codes the component parts with numbers.

Qualitative Measurement
Ways of measuring that systematically observes attitudes, beliefs, values and behaviors in their entirety.

Interview Guide
A list of the topics or questions that a researcher wants to raise in an in-depth interview.

Observation Guide
A list of the types of individual behavior or group interactions that a researcher wants to observe in an observation study.

.
Group Cohesiveness
How strongly group members are connected to other group members.

Boundary Maintenance
A social point that demarks the borderline between one or more groups.

Anomie
A state of normlessness.

Inter-rater Reliability
The condition in which two or more researchers using the same measures with the same study participants get the same results.

Constants
Measures to which responses don't vary. All responses are the same.

Chapter 2.7: Measurement

Closed-ended question	A question that has a defined set of mutually exclusive responses.
Mutually Exclusive	When the values of a measure do not overlap.
Open-ended question	A question without a defined set of responses for a respondent to choose from.
Index	Using multiple indicators to measure abstract or complicated concepts, such as quality of life. The indicators and their responses are grouped, often summed, together to create a more valid measure.
Likert Scale	A response set that reflects a continuum of agreement with a statement about the social world. For example, "How satisfied are you with your college education? Strongly Dissatisfied, Dissatisfied, Satisfied, Strongly Satisfied."
Levels of Measurement	Levels of categorization in quantitative measurement. They determine how variables can be analyzed.
Nominal	A response set with values that can't be ranked or ordered, such as, "Male or "Female."
Ordinal	A response set with values that can be ranked or ordered, such as, "Strongly Disagree, Disagree, Agree, Strongly Agree."
Interval	A response set with meaningful numerical values, such as age in years, income, or years of education.

TOPICS

One of the key things that we have learned in science is that we need to be able to measure something in order to study it. This is true whether you're doing biology, chemistry, psychology, economics or sociology. Of course applied sociology is no different. If you can't measure a problem, then you can't determine whether the interventions or programs you built to solve the problem worked! Hence, **measurement** is critical to applying sociology.

Let's look at an example. Suppose a travel company hires you to determine how satisfied people are with their vacations. The company wants to know this so that they can recommend enjoyable vacations to all of their clients. Seems easy. But how are you going to determine people's satisfaction? That is measurement. We need to develop a way to measure people's

Chapter 2.7: Measurement

satisfaction with their vacations. We could ask them, "Would you recommend this vacation to your friends and family? Yes or No." Or, we might ask, "How satisfied were you with your vacation? Are you strongly satisfied, somewhat satisfied, somewhat dissatisfied, or strongly dissatisfied?" Or we might go into the company's files and see what vacations people take repeatedly. Taking the same vacation more than once surely serves as an **indicator** of satisfaction.

Additionally, sociologists often want to transfer or **code** the answers to questions into numbers. For example, the responses "strongly dissatisfied" might be coded 1, "somewhat dissatisfied" might be coded 2, "somewhat satisfied" might be coded 3, and "strongly satisfied" might be coded 4. These numbers then become the data, not the written words. Sociologists use the numbered data to do statistical analyses, which we will discuss further in the next chapter. This idea of turning observations of social reality into numbers may be new to you, but it is basically the same thing as turning observations of physical reality into numbers, which you have been doing since you could talk. How many times have you answered the question, "How old are you?" That number is nothing other than a way to measure physical reality or the passing of time. People created that **measure** and gave it meaning over time. Same goes for the answers to "How tall are you?", "How much do you weigh?", "How much does it cost?", "What time is it?" and so on.

Not unlike the physical world, we can measure the social world. We call the tools that we create to measure the social and physical world **measurement instruments**. In the physical sciences, they have more exact measurement instruments, such as thermometers to measure temperature, scales to measure weight, or x-rays to measure cell growth. For the most part, they are not measuring things as slippery as attitudes, values, and beliefs. Nor do they have to rely on their research subject's to verbalize the information for which they are trying to measure. For example, a physician may ask you how you are feeling, but then she will also likely run a battery of lab tests to measure that same thing. And it is the latter she will rely on to diagnose and treat you. Sociologists do not have that luxury. We must design measures that accurately reflect the concept at hand, and questions that people can and will answer.

This leads us to the issues of **reliability** and **validity**. We want to develop measures and measurement instruments that are both reliable and valid. If a measure is reliable then it measures the same way every time. For example, let' say an exercise facility in a retirement community wants to evaluate the effectiveness of their program. After six months of participation, we might develop a telephone survey that, among other questions, will ask participants, "On a **scale** from 1 to 10, how much better do you feel today than you did six months ago?" Let's say we **pre-test** the survey by calling a few participants, one of which is Jane Doe. We ask Jane the question and she replies "7." Then we call her again the next day and ask her the same question and she replies "3." If the question was worded well then she should have responded similarly to the previous day, provided nothing drastic changed in her health after one day. So, in this case, we have evidence that the measure is unreliable. This could be due to the **question stem** or the **response set**. Possibly, the scale 1-10 did not have meaning to the respondent. Let's say we decide it is the response set, and change the response options to "A great deal better, somewhat better, the same, somewhat worse, and a

Chapter 2.7: Measurement

great deal worse." We pre-test again, and find that people answer the same upon asking them the question repeatedly across two to three days. The measure is now reliable.

If an instrument is valid then it actually measures what it is supposed to. For example, suppose we were trying to assess whether victimization is on the rise among a county's residents. The county wants this information to decide whether to budget additional monies for a crime prevention program. We develop a telephone survey, and one of the questions asks, "In the calendar year 1999, were you victimized? Yes, No or Don't Know." We pre-test the survey and find a large percentage of "Don't Know" responses, suggesting that people do not know what it means to be victimized. Many simply respond "no" because they figure if they don't know what it means than it must not have happened to them. The question is not valid. It is not measuring what we wanted it to measure. It may still be reliable. We could ask the same people the same question several times and still get a "no" response among people who were victimized because they consistently don't understand the question. Hence, a measure can be reliable but not valid. We decide to rewrite the question into ten questions that ask about specific types of victimization, such as "Did a person enter your home or car without your permission? Or, "Did anyone hit you?" People understand these questions better, and are able to respond accurately.

Finally before we try to actually create measures, we need to consider one more issue. The measurement issues we've discussed above largely pertain to **quantitative measurement**. Applied sociologists also use **qualitative measurement**. We've explained the difference between quantitative and qualitative approaches earlier in the book. Here we just want to remind you that qualitative and quantitative approaches to measurement differ. Qualitative measurement jumps directly into the social world, allowing the researcher to be the measurement instrument through his or her interactions in and observations of the social interaction. In this sense, qualitative measurement bypasses the need to create a standardized measurement instrument with questions and responses that will provide an accurate window on the social world. Instead, with qualitative measurement, researchers create an **interview or observation guide**.

An interview guide lists the topics or questions that you want to raise when you talk with people in in-depth interviews. Similarly, an observation guide lists the types of individual behavior or group interactions that you want to observe.   Each guide is supposed to do just that -- guide, not dictate, the questions asked or observations sought. When using qualitative measurement, the researcher has the flexibility to ask other questions that may arise from the discussion, or look for other interactions as suggested by whatever he or she observes. Let's look at an example of what you would put on interview and observation guides.

Suppose a local child day care center wants to improve how children interact with other children in their center. They have noticed that when new children enter a classroom, the children jockey for attention from each other and the teacher. You decide you will need to both observe the children's interaction, and then interview some of the children who interact well when new children join the class. On your observation guide, you create sections for the different group characteristics, attitudes, values, beliefs and behaviors that are salient to this project. Two

Chapter 2.7: Measurement

might be **group cohesiveness** and **boundary maintenance**. Underneath each section you list some specific behaviors and interactions that you might observe that would indicate each respective group characteristic. For example, underneath group cohesiveness you might list "children sharing/not sharing toys" or "nicknames." Underneath boundary maintenance you might list "inviting/rejecting others to join a game" or "existence of sub-groups or cliques." You get the idea. Then while you observe the children's interaction or soon afterwards, you write down examples of whatever is on your observation guide. And you revise the guide as you go, as is indicated by your observations.

An interview guide would follow the same process. But instead of listing behaviors or interactions, you list questions that you would like to ask the children, under their respective headings. For example, you might have a section on family life, the day care center, and new classmates. Then under each section you list specific topics or questions. Under family life, you might list "number of siblings," "position in sibling order," "parent's marital status and relations," and "a regular day at home." Discussion of these topics would indicate the quality of the child's home life, and their experience handling conflict or change. You would have similar kinds of topics and questions listed under the other sections. In observing all the children and interviewing the children who consistently play well, you might find that children without siblings or children who are the youngest in their families are primarily the children expressing difficulty when new children join the class. These children might not have the same skills in handling change as children who have younger siblings. When a new child joins the class, they perceive a state of **anomie**, and struggle to create a new social order. During this struggle, they act out. One of your recommendations might be that the day care center should routinely play games where the children rotate onto different teams so that they learn to deal with change and work with others.

Just like with quantitative measurement, reliability and validity issues arise with qualitative measurement. One of the most critical issues is what sociologists call **inter-rater reliability**. This term refers to the idea that how one person interacts in and observes the social world may vary dramatically from how another person would. In a qualitative study, if two or more interviewers or observers collect data on the same topic, setting and participants, they may not see, hear or note the same occurrences. The findings of such a study may not be easily reproduced, which is one of the hallmarks of science. If findings can't be reproduced, then the findings may not be reliable or valid. Interview and observation guides help to reduce this type of variation in interviewing and observation.

Also improving the reliability and validity, when an applied sociologist uses qualitative measurement, he or she looks for a pattern to reproduce several times across their observations and interviews. In sum, qualitative approaches to measurement can be a very powerful way to understand the social world we live in. They provide rich data, but usually take much longer to collect. Qualitative approaches are most useful when 1) the issue or problem being studied is too complicated to be translated into a series of standardized questions, 2) when not enough is known about the topic in order to develop standardized questions, or 3) when you can not access the target population with other methods.

Chapter 2.7: Measurement

TOOLS

You can see that creating measurement instruments is pretty creative stuff! But, measurement can get a bit tricky too. So, let's look at some measurement concepts that will help you develop measurement instruments. Remember, though, you must start any project by clearly identifying the problem and related issues. Without that step, measurement tools will not help. And keep in mind that the more complex the individual or group characteristic, attitude, value, belief or behavior you are studying, the more difficult it is to measure.

Tool	What about it?
Group vs. Individual Characteristics	In general, group characteristics are harder to measure than individual characteristics. Take, for example, group cohesiveness from the example above. Group cohesiveness requires more sophisticated measurement than, say, the individual characteristics of education, race, and religion.
Attitude, Value and Belief vs. Behavioral Questions	Behavior is often easier to measure than is an attitude, value or belief. We can look a data previously collected about people's behaviors, such as their criminal record, health care services received, or whether someone took a vacation at the same place twice. Or, we can observe a particular behavior happening in real time, such as how children play together. And, we can ask people about their behavior, such as "Have you ever smoked marijuana?" or "Do you usually vote democratic, republican, or for third party candidates? If the question is written well and the issue salient, people can generally recall past behavior.
Variables vs. **Constants**	You need to make sure that your measures will have some variation in response across individuals or groups. If not, you have a constant, which may complicate your statistical analyses. Often constants occur because a question did not provide responses that reflect the true variation in the population. For example, you ask fraternity members the number of alcoholic drinks they consumed per week and provide the following responses: 0, 1-2, 3-5, 6 or more. All fraternity members respond "6 or more." Your categories do not reflect the true variation. You need to break 6 or more into multiple categories.

Chapter 2.7: Measurement

Tool	What about it?
Closed-ended Question	A question that has a defined set of **mutually exclusive** responses, such as: What is your sex: 1) Female 2) Male Is your best friend a: 1) Woman, 2) Man, 3) Neither, 4) I don't know, 5) I don't have a best friend. What is your current marital status? 1) Married, 2) Separated, 3) Divorced, 4) Widowed, 5) Single.
Open-ended Question	A question without a defined set of responses, such as: What was your first job? _____ Sometimes the recorded responses are then organized into quantitative categories, such as different occupational categories.
Index	Some concepts require multiple indicators because they are very complicated or abstract. Using multiple indicators to measure a complicated concept is a form of triangulation, giving us a better view of reality. For example, you can't measure quality of life with one measure. It has several dimensions to it, such as crime rates, pollution, unemployment rate, cost of living, quality of educational institutions, health care services, parks and recreational amenities, weather, etc… If we wanted to measure quality of life, we would create a measure for each of these dimensions. Then, to summarize quality of life, we would sum each respondent's answer to all of the questions; thereby creating an additive index.

Chapter 2.7: Measurement

Tool	What about it?
Likert Scale	A Likert scale is a response set that reflects a continuum of agreement with a statement about the social world. For example, to what extent do you agree or disagree with abortion? 1. Strongly Disagree 2. Disagree 3. Undecided 4. Agree 5. Strongly Agree Sometimes the agree-disagree labels do not fit the issue at hand. In this case, applied sociologists would use labels such as "Frequently," "Some," "Rarely," "Never," or "Very Unlikely," "Somewhat Unlikely," "Somewhat Likely," "Very Likely." Applied sociologists may omit the middle or neutral category if all respondents should have an opinion on the issue at hand.

One more thing. Before you start creating measures you need to decide how you will want to analyze your data. As we stated earlier, we will discuss data analysis in the next chapter. For now, just take our word for it that the level at which the social world is measured directly determines what types of analyses you can do. So you need to consider your analysis needs while developing your measures. There are four **levels of measurement**: **nominal**, **ordinal**, **interval**, and **ratio**. Each of these levels of measurement has certain properties to them, and can serve as tools to guide the measures you develop. The table on the following page quickly explains each level of measurement.

Chapter 2.7: Measurement

Level of Measurement	What is it?	Examples
Nominal	Measures with categorized responses that have no inherent order.	Yes No Male Female African American, American Indian, Asian, Eastern Indian, Latino, White
Ordinal	Measures with categorized responses in an ordered list. Categories can be ranked from highest to lowest, but the social distance between consecutive categories is not equal.	Strongly Disagree, Disagree, Not Sure, Agree, Strongly Agree Less than High School, High School Diploma, Some College, College Degree
Interval	Measures with meaningful numeric responses. The distance between consecutive categories or numbers is the same.	1) What is your age: _____ 2) Use the scale below to indicate how happy you are: Not happy Extremely At all Happy 0 1 2 3 4 5 6 7 8 9 3) How many years of education have you completed? _____

TASKS

1. The US Marine Corp hires your company to help them develop programs to improve their soldier's marital relationships. Their data suggests that their soldiers have one of the lowest rates of marital satisfaction in the country, and that this is one of the main reasons why soldiers leave the Marines. You developed a program and implemented it. Now you must determine whether the program you developed is working. To do so, you must determine whether couples are more satisfied with their relationship now?

   - How could you measure marital satisfaction?
   - How could you measure marital satisfaction in an in-depth interview with each spouse?

Chapter 2.7: Measurement

- How could you measure marital satisfaction through observing couples interactions?

2. You work for a city manager. A group of city employees are picketing over domestic partnership rights in front of the city hall in a suburb of Burlington, Vermont. They want the city to extend health insurance benefits to the romantic partners of gay and lesbian city employees. Recently, the State of Vermont extended marriage rights to gay and lesbian couples. These employees argue that, given that the state now recognizes gay and lesbian marriages, that the legal partners of gay and lesbian employees should be treated equally to the legal partners of heterosexual employees. The mayor and the city manager want to determine how much support there is for domestic partnership rights among the city's residents.

- Identify what main data collection and sampling method(s) you will use.
- Determine what topics for which you will need to write questions.
- Develop questions and question stems to measure those topics.
- Identify any potential problems with the way your measurements.