

Chapter Outlines for:

Frey, L., Botan, C., & Kreps, G. (1999). *Investigating communication: An introduction to research methods*. (2nd ed.) Boston: Allyn & Bacon.

Chapter 5: Designing Valid Communication Research

I. Introduction

A. People frequently refer to the concept of validity in everyday conversation, saying things like, “She has a valid point” or “That’s not a valid statement.”

1. They are referring to the accuracy of a statement; therefore, the best synonym for the word **validity** is *accuracy*.

II. Internal and External Validity

A. Two general types of validity are important: internal and external.

1. **Internal validity** concerns the *accuracy of conclusions* drawn from a particular research study.

a. Internal validity asks whether a research study is designed and conducted such that it leads to accurate findings about the phenomena being investigated for the particular group of people or texts being studied.

2. **External validity** concerns the *generalizability of the findings* from a research study.

a. External validity asks whether the conclusions from a particular study can be applied to other people/texts, places, and/or times.

3. The best studies are high on both internal and external validity.

a. There are times when a researcher must sacrifice a little of one type of validity to boost the other.

b. Validity ranges on a continuum from studies that have internal and/or external validity to those with less.

B. As Figure 5.1 shows, internal validity is potentially compromised by three general threats, each of which contains several specific threats:

1. How the research is conducted, effects due to research participants, and/or effects due to researchers.

2. Three factors may influence external validity and include: How the people/texts studied were selected, called *sampling*; whether the procedures used mirror real life, called *ecological validity*; and/or the need to *replicate* the findings.

III. Measurement Validity and Reliability

A. Data collected through questionnaires, interviews, and observations are worthwhile only if they are recorded in accurate ways.

1. **Measurement validity** refers to how well researchers measure what they intend to measure.

a. Measurement validity refers to the ability of a measurement technique to tap the referents of the concepts being investigated.

b. For any measurement to be valid, it must first be **reliable**, a term that implies both consistency and stability.

i. **Measurement reliability** means measuring something in a consistent and stable manner.

ii. Measurement validity and reliability go hand in hand; neither is meaningful without the other.

iii. A measurement can be reliable, however, but not necessarily valid.

2. **Measurement reliability** is not an absolute measure as indicated by the “friend who usually arrives on time”; there is some variation.

a. Even reliable or general consistent measurements have some deviation, or error.

b. Observed Measurement = True Score + Error Score which also shape the amount of

random error and measurement error (See the text's formula).

- i. **True score component:** In theory, the true score as measured over the course of time.
 - ii. **Error score component:** Is the amount of deviation from the true value.
 - iii. **Random error** (also called **accidental or chance error**): This is error that cannot be predicted or controlled.
 - iv. **Measurement error:** This is error that occurs due to faulty measurement procedures on the times the behavior is measured and, therefore, is more directly under the researchers' control.
 - (a) Measurement error can be reduced in several ways, such as conducting a **pilot study**; that is, a preliminary study that tests the questionnaires, interview protocols, observational techniques, and other methods to make sure they are as effective as possible before the main study starts.
- c. A perfectly reliable measurement would be 100 percent reliable.
- i. The reliability of most measurements ranges between 0 percent and 100 percent.
 - ii. A **reliability coefficient** provides a numerical indicator that tells the percentage of time a measurement is reliable.
 - iii. Reliability coefficients range from .00 (no consistency) to 1.00 (perfect consistency; 100 percent).
 - iv. Several techniques are used to assess the reliability of measurements including: comparing the results from *multiple administrations* of a measurement procedure; comparing the internal consistency of a measurement procedure administered once, called *single administration techniques*; and/or the amount of *agreement between observers* to assess observational measurement techniques.
3. Multiple administration techniques involve assessing the temporal stability of instruments at two or more points in time.
- a. **Test-retest method:** Administers the *same* measurement procedure to the same group of people at different times.
 - b. The measurement is considered reliable if the results are consistent (usually .70 or greater; called the *coefficient of stability*) from one time to another.
 - i. Just because a measurement varies from one time to another does not necessarily mean it isn't reliable.
 - ii. **Alternative procedure method:** Involves having the same people complete another, equivalent instrument at the second administration.
 - (a) **Coefficient of equivalence:** Statistically comparing scores on these two instruments is the basis for making claims about the reliability of the first instrument.
4. Most measurements are only taken once and involve single administration techniques to assess reliability.
- a. In single administrations, reliability is assessed by measuring the **internal consistency** of an instrument.
 - i. Because most instruments are composed of items measuring the same general concept, researchers can examine whether there is sufficient consistency between how people answer the related items.
 - ii. **Split-half reliability:** Assessed by separating people's answers on an instrument into two parts (half the questions in one part and half in the other) and then comparing the two halves.
 - (a) *First-half/last-half split* divides the answers at the midway point of the scale.
 - (b) Figure 5.2 demonstrates a better technique by showing how to calculate the *odd-even halves* reliability coefficient.
 - © Researchers can also create *balanced halves* if they are concerned that an item

characteristic might make a difference; or *random halves* such that items have an equal chance of being assigned to either half.

(d) One of the most popular internal consistency methods is Cronbach's (1951) **alpha coefficient method**, which assesses the overall relationship among the answers as the reliability coefficient for an instrument.

5. When a measurement instrument is found to not be reliable enough, researchers can measure the relation of each single item to the rest of the instrument, called the **agreement coefficient**, and try dropping problematic items.
6. In observational research, the most common method for assessing reliability is calculating the percentage of agreement between or among the observations of independent coders, which is called **interobserver, interrater, or intercoder reliability**.
 - a. If the observations recorded by two or more individuals who are not aware of the purposes of the study and each other's codings are highly related (showing 70 percent agreement or more), their ratings are considered reliable.
 - b. If the interobserver reliability score is lower than .70, several options are available ranging from modifying the observational system to increasing the length or number of observations sessions until consistent data are obtained.

B. Measurement Validity

1. There is an important difference between measurement reliability and measurement validity with the former being assessed numerically and the latter being assessed conceptually.
2. Content validity: A measurement instrument, such as a questionnaire, possesses **content validity** if it measures the attributes of the content being investigated.
 - a. The identification of these attributes ranges from relative ease to great difficulty.
 - b. One way a researcher can establish content validity is to establish **face validity** by generating items that "on the face of it," seem to accurately reflect the concept being investigated.
 - c. A stronger approach for establishing content validity is a **panel approach**, where qualified people are recruited to describe the aspects of that variable or to agree that an instrument taps the concept being measured.
3. **Criterion-related validity**: Established when a measurement technique is shown to relate to another instrument or behavior (called the criterion) already known to be valid.
 - a. The first type of criterion-related validity is called, **concurrent validity (also called convergent validity)** is established when the results from a new instrument agree with those from an existing, known-to-be valid criterion.
 - i. Concurrent validity can also be established by seeing how a group of experts (called a *criterion group*) perform with respect to a measurement instrument.
 - b. **Predictive validity**: Refers to how well a measurement instrument forecasts or predicts an outcome.
4. **Construct validity** is the extent to which scores on a measurement instrument are related to in logical ways to other established measures.
5. Triangulating validity: Recognizes that content, criterion-related, and construct validity are not mutually exclusive.
 - a. The best way researchers can argue for the validity of a measurement technique is by trying to establish all three types.
 - b. Designing reliable and valid measurement instruments is a crucial part of research that does not come easy.
 - i. Researchers must engage in a systematic process of constructing and testing instruments (See the guidelines reported in Figure 5.3).

IV. Threats to Internal Validity

- A. Besides measurement validity, many other threats to internal validity may affect the

accuracy of the results obtained from a research study.

- B. These threats fall into three interrelated categories: threats due to *how research is conducted*, threats due to *research participants*, and threats due to *researchers*.
- C. A number of threats are due to the internal validity of a study are due to how research is conducted.
1. **Procedure validity and reliability:** Conducting research by administering accurate measurement techniques in a consistent manner.
 - a. **Treatment validity and reliability:** Making sure that any treatment a study is investigating is what it purports to be every time it is administered.
 - i. To make sure treatments are valid and reliable, researchers conduct **manipulation checks** in which some of the respondents are exposed to a study's independent variable(s), while others in a control group are not exposed.
 - b. **Controlling for environmental influences:** Keeping the setting in which a study is done as consistent as possible.
 2. **History** refers to changes in the environment external to a study that influence people's behavior within the study.
 - a. The threat of history is particularly important for longitudinal research that follows people over a relatively long period of time.
 3. The **sleeper effect** refers to an effect that is not immediately apparent but becomes evidenced over the course of time.
 4. **Sensitization** (sometimes called **testing, practice effects, pretest sensitizing**) is the tendency for an initial measurement in a research study to influence a subsequent measurement.
 - a. Sensitization is a particularly significant problem that must be ruled out in studies of instructional behavior.
 5. **Data analysis:** An important threat to the internal validity of research has to do with the way in which data are analyzed.
 - a. Researchers sometimes use improper procedures to analyze data, and this may lead to invalid conclusions.
- D. A number of threats are due to the internal validity of a study are due to research participants.
1. **Hawthorne effect:** Any change in behavior due primarily to the fact that people know they are being observed.
 - a. The Hawthorne effect is derived from a famous study about the effects of the amount of light in a room on worker productivity.
 - i. Workers produced more not due to the changes in lighting, but due to knowing that they were being observed.
 - b. To the extent that people engage in atypical behavior because they know they are being observed, conclusions about their behavior may not be valid.
 2. **Selection** of people or texts for a study may influence the validity of the conclusions drawn.
 - a. A particular type of selection problem is **self-selection bias**, which can occur when researchers compare groups of people that have been formed on the basis of self-selection.
 - i. One of the more insidious problems with selection is that it interacts with many of the other internal validity threats to produce unique effects.
 3. Another way in which selection can threaten the validity of research findings is **statistical regression** (also called **regression toward the mean**), the tendency for individuals or groups selected on the basis of initial extreme scores on a measurement instrument to behave less atypically the second and subsequent times on that same instrument.
 - a. Such an effect is also called a **regression effect** or **artifact**.

- b. Choosing people on the basis of extreme scores also potentially leads to ceiling and floor effects.
 - i. A **ceiling effect occurs** when people have scores that are at the upper limit of a variable, making it difficult to tell whether a treatment has any effect.
 - ii A **floor effect** occurs when people have scores that are at the lower limit of a variable, making it difficult to tell whether any treatment has an effect.
 - 4. **Mortality** (also called **attrition**) is the loss of research participants from the beginning to the end of a research study.
 - 5. **Maturation** refers to internal changes that occur *within* people over the course of a study that explains behavior.
 - a. These changes can be physical or psychological and can happen over a short or long time period.
 - 6. **Interparticipant bias** or **intersubject bias** results when the people being studied influence one another.
 - a. If the people being studied, converse before or during the study, and the result is that the experimental treatment becomes known, a **diffusion of treatments** or **diffusion effect** will occur.
- E. A number of threats to the internal validity of a study are due to researchers, or **researcher effects**, which refer to the influence of researchers on the people being studied.
1. The **researcher personal attribute effect** occurs when particular characteristics of a researcher influence people's behavior.
 - a. This effect is likely to occur when the research task is ambiguous, and participants look to the researcher for clues about how to behave.
 - b. This effect is more likely to occur when the task is related to the personal characteristics of a researcher.
 2. The **researcher unintentional expectancy effect** (also called the **Rosenthal** or **Pygmalion effect**) occurs when researchers influence research participants' by inadvertently letting them know the behavior they desire.
 - a. Researchers may smile unconsciously when participants behave in ways that confirm their hypothesis or frown when they behave in ways that don't support the hypothesis.
 - b. Such behavior is often referred to as *demand characteristics*.
 - c. To control for this potential threat, some researchers go to great lengths to remove themselves from the actual conducting of study.
 3. **Researcher observational biases** occur whenever researchers, their assistants, or the people they employ to observe research participants demonstrate inaccuracies during the observational process.
 - a. **Observational drift** occurs when observers become inconsistent in the criteria used to make and record observations.
 - b. A more insidious bias, **observer bias**, occurs when the observers' knowledge of the research (e.g., its purpose or hypotheses) influences their observations.
 - c. The **halo effect** occurs when observers make multiple judgements over time and typically overrate a research participant's performance because that participant did well (or poorly) in an earlier rating.
- V. External Validity
- A. In addition to knowing that conclusions drawn are valid for the people/texts studied (internally valid), researchers also want to be able to generalize these conclusions to other people/texts, and times.
 - B. Judgments of external validity, the ability to generalize findings from a study to others, are made on the basis of three issues: sampling, ecological validity, and replication.
 1. Communication researchers are interested in a **population** (also called a **universe** when applied to texts) of communicators, all the people who possess a particular

characteristic, or in the case of those who study texts, all the messages that share a characteristic of interest.

- a. The population of interest to researchers is often called the **target group**.
- b. The best way to generalize to a population is to study every member of it, or what is called a **census**.
 - i. Communication researchers are often interested in large populations where obtaining a census is practically impossible.
- c. Because measuring every member of a population usually isn't feasible, most researchers employ a **sample**, a subgroup of the population.
 - i. The results from a sample are then generalized back to the population.
 - ii. For such a generalization to be valid (demonstrate **population validity**), the sample must be *representative* of its population; that is, it must accurately approximate the population.
- d. **Random sampling** (also called **probability sampling**) involves selecting a sample in such a way that each person in the population of interest has an equal chance of being included.
 - i. No random sample ever represents perfectly the population from which it is drawn, and **sampling error** is a number that expresses how much the characteristics of a sample differ from the characteristics of a population.
 - ii. **Simple random sample**: Each person of the population is assigned a consecutive number and then selected from these numbers in such a way that each number has an equal chance of being chosen.
 - (a) To conduct a simple random sample, a researcher must first have a complete list of the population.
 - (b) Use a formal procedure that guarantees that each person in the population has an equal chance of being selected.
 - © One strategy is to use a **random number table** (See Appendix A) that lists numbers generated by a computer in a nonpurposive way, which means there is no predetermined relationship whatsoever among the numbers on the table.
 - (d) *Sampling without replacement* means if you come across the same number twice, skip it.
 - (e) *Sampling with replacement* means replacing the number each time so that it has another chance of being selected.
 - (f) Communication researchers often use a simple random sampling procedure because it is an easy way of drawing a relatively representative sample from a population.
 - (g) The biggest difficulty in conducting a simple random sample is obtaining a complete list of a population.
- e. A **systematic sample** (also called **ordinal sampling**) chooses every *n*th person from a complete list of a population *after starting at a random point*.
 - i. The interval used to choose every *n*th person is called **sampling rate**.
 - ii. Systematic samples are often used with very large populations, perhaps because they are easier to employ than a simple random sample.
 - iii. Systematic random samples usually produce virtually identical results to those obtained from a simple random sample.
- f. A **stratified sample** categorizes a population with respect to a characteristic a researcher concludes to be important, called a **stratification variable**, and then samples randomly from each category.
 - i. Stratified samples are used frequently in research, especially in political polling, because they ensure that different subgroups of a population (a subgroup of a population is called a **stratum**; the plural is **strata**) are represented in a sample.

- ii. **Proportional stratified random sample:** Respondents from stratified categories can be randomly selected in proportion to their representation in the population.
- iii. The frequency with which respondents can be obtained from stratified populations is called the **incidence**.
- g. Each of the preceding types of samples necessitates obtaining a complete list of the population of interest and then randomly selecting members from it.
 - i. As noted previously, this is not always feasible; however, in the case of IBM for example, it would be possible to compile a list of all branch offices and then randomly select the units to be studied; a **cluster sample**.
 - ii. Many cluster samples involve **multistage sampling**, a procedure for selecting a sample in two or more stages, in which successively smaller clusters are picked randomly.
 - iii. Multistage sampling is particularly appropriate for companies that conduct state, regional, or national surveys, such as the Gallup polls or Nielsen ratings.
- h. It sometimes isn't possible to sample randomly from a population because neither a complete participation list nor a list of clusters is available.
 - i. When random samples cannot be obtained, **nonrandom sampling** (also called **nonprobability sampling**) is used and consists of whatever researchers do instead of using procedures that ensure each member of a population has an equal chance of being selected.
- i. In a **convenience sample** (also called an **accidental or haphazard sample**), respondents are selected nonrandomly on the basis of availability.
 - i. The most popular type of convenience sample for researchers who teach at universities is one composed of students.
 - ii. The problem with convenience samples, as with all nonrandom samples, is that there is no guarantee the respondents chosen are similar to the general population they are supposed to represent.
 - iii. Sometimes researchers can increase that confidence by making the convenience sample similar in some ways to the population that respondents are supposed to represent.
- j. In a **volunteer sample**, respondents choose to participate in a study.
 - i. To recruit volunteers, researchers frequently offer some reward to those people who sign up for studies, especially to university students.
 - ii. Researchers must be very wary about generalizing the results from a volunteer sample to the population of interest.
- k. In a **purposive sample** (also called a **deliberate, purposeful, or strategic sample**), respondents are nonrandomly selected on the basis of a particular characteristic.
 - i. A purposive sample is similar to a stratification (random) sample in that the characteristic chosen is a stratification variable; yet, the crucial difference is that there is no random selection in purposive samples.
- l. To gather a **quota sample**, respondents are selected nonrandomly on the basis of their known proportion in a population.
- m. In a **network sample** (also called a **multiplicity sample**), respondents are asked to refer a researcher to other respondents, and these people are contacted, included in the sample, and, in turn, are asked for the names of additional respondents.
 - i. This type of technique is sometimes called the **snowball technique**, as a network sample becomes larger as a researcher contacts more people.
 - ii. The snowball technique is quite popular with naturalistic researchers (see Chapter 10).

VI. Ecological Validity

A. **Ecological validity** refers to research that describes what actually occurs in real-life

circumstances.

- B. To the extent that research procedures reflect what people do in the contexts in which their behavior normally occurs, confidence in the generalizability of the findings to other people and situations is increased.
1. One way to increase the ecological validity of communication research is to study message behavior as it occurs in natural settings (see Chapter 10).
 2. Studying communication behavior in natural settings increases the generalizability of research because communication processes may be thought of as streams of behavior.
 3. A critical reader must ask serious questions about the validity of findings from research conducted in controlled settings.
 4. This is not to suggest that all research studies conducted in natural settings are more ecologically valid than research conducted in laboratories.
 5. One good way to enhance the generalizability of laboratory findings is to replicate them in natural settings.
 - a. Sometimes, **confederates** (persons who pretend to be research participants to help a researcher) are used in such replication attempts.
- C. The third criterion in establishing external validity, **replication**, refers to conducting a study that repeats or duplicates in some systematic manner a previous study.
1. There is no way to ever replicate someone else's study exactly, and there are three types of replication.
 - a. An **exact replication** duplicates a research study as closely as possible, with the exception of studying different research participants.
 - i. Exact replications in the social sciences are rare.
 2. A **partial replication** duplicates a previous research study by changing one procedure while keeping the rest of the procedures the same.
 - a. Partial replications are the most frequent type employed, for they typically are designed to both retest and extend the findings from a previous study.
 - b. One replication does not prove an alternative model.
 3. A **conceptual replication** examines the same issue as a previous study but uses entirely different procedures, measurement instruments, sampling procedures, and/or data-analytic techniques.
 - a. The goal is to see if the same results can be obtained with very different research procedures.
 4. While replication is tremendously important, not all studies are worth replicating and not all replications are equivalent.
 - a. Rosenthal proposes three criteria for judging the value of a replication: *When*, *how*, and *by whom* the replication is conducted.
 - i. He argues that earlier replications are more valuable than later ones.
 - ii. He argues that how similar the methodology of a replication is to the original study and how similar the findings are determines the type of information gained.
 - iii. He contends that replications by different researchers are more valuable than replications by the same researcher, and that the more the replication is different from the previous researcher, the better.
 - b. In addition to the preceding criteria, the number of replications is important to consider and in general, the more the better, such as three are better than two; however, there is also point of diminishing returns after many replications have been done.
 5. There is healthy regard and respect for replication in the physical sciences, where it is a normal part of the way researchers do business.
 - a. In the social sciences, such as Communication, replication is viewed in contradictory ways; with some seeing it as a natural part of "science," while others such as many editorial boards of scholarly journals do not see it as necessary or beneficial.
 - b. The danger in not replicating research is that invalid results from a single study may

end up guiding behavior.

VII. Conclusion

- A. The importance of validity cannot be overestimated.
- B. Researchers and consumers must invest the efforts and diligence described in this chapter to produce and evaluate findings that can be trusted.