Chapter 8
Validity of Research Results

(Reminder: Don't forget to utilize the concept maps and study questions as you study this and the other chapters.)

In this chapter we discuss validity issues for quantitative research and for qualitative research.

Validity Issues in the Design of Quantitative Research

On page 228 we make a distinction between an extraneous variable and a confounding variable.

- An extraneous variable is a variable that **MAY compete** with the independent variable in explaining the outcome of a study.
- A confounding variable (also called a third variable) is an extraneous variable that **DOES** cause a problem because we know that it **DOES** have a relationship with the independent and dependent variables. A confounding variable is a variable that systematically varies or influences the independent variable and also influences the dependent variable.
- When you design a research study in which you want to make a statement about cause and effect, you must think about what extraneous variables are probably confounding variables and do something about it.
- We gave an example of "The Pepsi Challenge" (on p. 228) and showed that anything that varies with the presentation of Coke or Pepsi is an extraneous variable that may confound the relationship (i.e., it may also be a confounding variable). For example, perhaps people are more likely to pick Pepsi over Coke if different letters are placed on the Pepsi and Coke cups (e.g., if Pepsi is served in cups with the letter "M" and Coke is served in cups with the letter "Q"). If this is true then the variable of cup letter (M versus Q) is a confounding variable.
- In short we must always worry about extraneous variables (especially confounding variables) when we are interested in conducting research that will allow us to make a conclusion about cause and effect.
- There are four major types of validity in quantitative research: statistical conclusion validity, internal validity, construct validity, and external validity. We will discuss each of these in this lecture.

Statistical Conclusion Validity

Statistical conclusion validity refers to the ability to make an accurate assessment about whether the independent and dependent variables are related and about the strength of that relationship. *So the two key questions here are 1) Are the variables related? and 2) How strong is the relationship?*

- Typically, null hypothesis significance testing (discussed in Chapter 16) is used to determine whether two variables are related in the population from which the study data were selected. This procedure will tell you whether a relationship is statistically significant or not.
- For now, just remember that a relationship is said to be statistically significant when we do NOT believe that it is nothing but a chance occurrence, and a

relationship is not statistically significant when the null hypothesis testing procedure says that any observed relationship is probably nothing more than normal sampling error or fluctuation.

- To determine how STRONG a relationship is, researchers use what are called effect size indicators. There are many different effect size indicators, but they all tell you how strong a relationship is.
- For now remember that the answer to the first key question (Are the variables related?) is answered using null hypothesis significance testing, and the answer to the second key question (How strong is the relationship?) is answered using an effect size indicator.
- The concepts of significance testing and effect size indicators are explained in Chapter 16.

Internal Validity

When I hear the term "internal validity" the word **cause** always comes into my mind. That's because internal validity is defined as the "approximate validity with which we infer that a relationship between two variables is causal" (Cook and Campbell, 1979. P.37).

- A good synonym for the term internal validity is ***causal validity*** because that is what internal validity is all about.
- If you can show that you have high internal validity (i.e., high causal validity) then you can conclude that you have strong evidence of causality; however, if you have low internal validity then you must conclude that you have little or no evidence of causality.

Types of Causal Relationships

There are two different types of causal relationships: causal description and causal explanation.

- Causal description involves describing the consequences of manipulating an independent variable.
- In general, causal description involves showing that changes in variable X (the IV) cause changes in variable Y (the DV): **X---->Y**
- Causal explanation involves more than just causal description. Causal explanation involves explaining the mechanisms through which and the conditions under which a causal relationship holds. This involves the inclusion (in your research study) of mediating or intervening variables and moderator variables. Mediating and moderator variables are defined in Chapter Two in Table 2.2 (on page 36). To see Table 2.2, click here.

Criteria for Inferring Causation

There are three main conditions that are always required if you want to make a claim that changes in one variable cause changes in another variable. We call these the **three necessary conditions for causality**.

- These three conditions are summarized below in Table 11.1:

TABLE 11.1
The Three Necessary Conditions for Causation

Researchers must establish three conditions if they are to conclude that changes in variable A cause changes in variable B.

Condition 1:	Variable A and variable B must be related (the relationship condition).
Condition 2:	Proper time order must be established (the temporal antecedence condition).
Condition 3:	The relationship between variable A and variable B must *not* be due to some confounding extraneous or "third" variable (the lack of alternative explanation condition).

- If you want to conclude that X causes Y you must make sure that the three above necessary conditions are met. It is also helpful if you have a theoretical rationale explaining the causal relationship.
- For example, there is a correlation between coffee drinking and likelihood of having a heart attack. One big problem with concluding that coffee drinking causes heart attacks is that cigarette smoking is related to *both* of these variables (i.e., we have a Condition 3 problem). In particular, people who drink little coffee are less likely to smoke cigarettes than are people who drink a lot of coffee. Therefore, perhaps the observed relationship between coffee drinking and heart attacks is the result of the extraneous variable of smoking. The researcher would have to "control for" smoking in order to determine if this rival explanation accounts for the original relationship.

Threats to Internal Validity

In this section, we discuss several threats to internal validity that have been identified by research methodologists (especially by Campbell and Stanley, 1963).

- These threats to internal validity usually call into question the third necessary condition for causality (i.e., the "lack of alternative explanation condition").

Before discussing the specific threats, I want you to get the basic idea of two **wea**k designs in your head.

- The first weak design is the one is the one-group pretest-posttest design which is depicted like this:

O X O

In this design, a group is pretested, then a treatment is administered, and then the people are post tested. For example, you could measure your students' understanding of history at the beginning of the term, then you teach them history for the term, and then you measure them again on their understanding of history at the end of the term.

- The second weak design to remember for this chapter is called the posttest-only design with nonequivalent groups. In this lecture, I will also refer to this design as a two-group design and sometimes as a multigroup design (since it has more than one group).

$$\begin{array}{ll} \mathbf{X_{Treatment}} & \mathbf{O_2} \\ \text{----------} & \text{----------} \\ \mathbf{X_{Control}} & \mathbf{O_2} \end{array}$$

In this design, there is no pretest, one group gets the treatment and the other group gets no treatment or some different treatment, and both groups are post tested (e.g., you teach two classes history for a quarter and measure their understanding at the end for comparison). Furthermore, the groups are found wherever they already exist (i.e., participants are not randomly assigned to these groups).

- In comparing the two designs just mentioned note that the comparison in the one group design is the participants' pretest scores with their posttest scores. The comparison in the two group design is between the two groups' posttest scores.
- Some researchers like to call the point of comparison the "counterfactual." In the one-group pretest- posttest design shown above the counterfactual is the pretest. In the two-group design shown above the counterfactual is the posttest of the control group.
- Remember this key point: In each of the multigroup research designs (designs that include more than one group of participants), you want the different groups to be the same on all extraneous variables and different ONLY on the independent variable (e.g., such that one group gets the treatment and the other group does not). In other words, you want the only *systematic* difference between the groups to be exposure to the independent variable.

The first threat to internal validity is called **ambiguous temporal precedence**.

- Ambiguous temporal precedence is defined as the inability of the researcher (based on the data) to specify which variable is the cause and which variable is the effect.
- If this threat is present then you are unable to meet the second of the three necessary conditions shown above in Table 11.1. That is, you cannot establish proper time order so you cannot make a conclusion of cause and effect.

The second threat to internal validity is called the **history** threat.

- The history threat refers to any event, *other than the planned treatment* event, that occurs between the pretest and posttest measurement and has an influence on the dependent variable.
- In short, if both a treatment and a history effect occur between the pretest and the posttest, you will not know whether the observed difference between the pretest and the posttest is due to the treatment or due to the history event. In short, these two events are confounded.

- For example, the principal may come into the experimental classroom during the research study which alters the outcome.
- The history effect is a threat for the one group design but it is **not** a threat for the multigroup group design.
- You probably want to know why this it true. Well, in the one group design (shown above) you take as your measure of the effect of the treatment the difference in the pretest and posttest scores. In this case, this all or part of the difference could be due to a history effect; therefore, you don't know whether the change in scores is due to the treatment or to the history effect. They are confounded.
- The basic history effect is n**ot a threat to the two group design** (shown above) because now you are comparing the treatment group to a comparison group, and as long as the history effect occurs for both groups the **difference** between the two groups will not be because of a history effect.

The third second threat to internal validity is called **maturation.**

- Maturation is present when a physical or mental change occurs over time and it affects the participants' performance on the dependent variable.
- For example, if you measure first grade students' ability to perform arithmetic problems at the beginning of the year and again at the end of the year, some of their improvement will probably be due to their natural maturation (and not just due to what you have taught them during the year). Therefore in the one group design, you will not know if their improvement is due to the teacher or if it is due to maturation.
- Maturation is not a threat in the two group design because as long as the people in **both** groups mature at the same rate, the **difference** between the two groups will **not** be due to maturation.

If you are following this logic about why these first two threats to internal validity are a problem for the **one group** design but not for the **two group** design then you have one of the major points of this chapter. This same logic is going to apply to the next three threats of testing, instrumentation, and regression artifacts.

The fourth threat to internal validity is called **testing.**

- Testing refers to any change on the second administration of a test as a result of having previously taken the test.
- For example, let's say that you have a treatment that you believe will cause students to reduce racial stereotyping. You use the one group design and you have your participants take a pretest and posttest measuring their agreement with certain racial stereotypes. The problem is that perhaps their scores on the posttest are the result of being sensitized to the issue of racial stereotypes because they took a pretest.
- Therefore in the one group design, you will not know if their improvement from pretest to posttest is due to your treatment or if it is due to a testing effect.
- Testing is not a threat in the two group design because as long as the people in **both** groups are affected equally by the pretest, the **difference** between the two

groups will **not** be due to testing. The two groups do differ on exposure to the treatment (i.e., one group gets the treatment and the other group does not).

The fifth threat to internal validity is called **instrumentation.**

- Instrumentation refers to any change that occurs in the way the dependent variable is measured in the research study.
- For example, let's say that one person does your pretest assessment of students' racial stereotyping but you have a different person do your posttest assessment of students' stereotyping. Also assume that the second person tends to overlook much stereotyping but that the first person picks up on all stereotyping. The problem is that perhaps much of the positive gain occurring from the pretest to the posttest is due to the posttest assessment not picking up on the use of stereotyping.
- Therefore in the one group design, you will not know if their improvement from pretest to posttest is due to your treatment for reducing stereotyping or if it is due to an instrumentation effect.
- Instrumentation is not a threat in the two group design because as long as the people in **both** groups are affected equally by the instrumentation effect, the **difference** between the two groups will **not** be due to instrumentation.

The sixth threat to internal validity is called **regression artifacts** (or **regression to the mean**).

- Regression artifacts refers to the tendency of very high pretest scores to become lower and for very low pretest scores to become higher on post testing.
- You should always be on the lookout for regression to the mean when you select participants based on extreme (very high or very low) test scores.
- For example, let's say that you select people who have extremely high scores on your racial stereotyping test. Some of these scores are probably artificially high because of transient factors and a lack of perfect reliability. Therefore, if stereotyping goes down from pretest to posttest, some or all of the change may be due to a regression artifact.
- Therefore, in the one group design you will not know if improvement from pretest to posttest is due to your treatment or if it is due to a regression artifact.
- Regression artifacts is not a threat in the two group design because as long as the people in **both** groups are affected equally by the statistical regression effect, the **difference** between the two groups will **not** be due to regression to the mean.

The seventh threat to internal validity is called **differential selection.**

- Differential selection only applies to multigroup designs. It refers to selecting participants for the various **groups** in your study that have **different** characteristics.
- Remember, we want our groups to be the same on all variables except the treatment variable; the treatment variable is the only variable that we want to be systematically different for the groups.

- Table 8.1 list a few of the may characteristics on which the students in the different groups may differ (e.g., age, anxiety, gender, intelligence, reading ability, etc.).
- Unlike the previous five threats, selection is **not an internal validity problem for the one group design but it is a problem for the two or multigroup design.**
- Looking at the definition again, you can see that selection is defined for two or multigroup designs. It is not relevant to the internal validity of the single group design.
- As an example, assume that you select two classes for your study on reducing racial stereotyping. You use two fifth grade classes as your groups. One group will get your treatment and the other will act as a control. The problem is that these two groups of students may differ on variables **other than** your treatment variable and any differences found at the posttest may be due to these "differential selection" differences rather than being due to your treatment.

The eight threat to internal validity is called **differential attrition** (it is also sometimes called mortality)**.** Attrition simply refers to participants dropping out of your research study.

- Differential attrition is the **differential** loss of participants from the various comparison groups.
- Just like the last threat, differential attrition is a problem for two or multigroup design but not for the single group design. (Notice the word differential in the definition.)
- For example, assume again that you are doing a study on racial stereotyping. Do you see how your result would be compromised if the kind of children that are most likely to have racial stereotypes drop out of **one** of your groups but not the other group? Obviously, the difference observed at the post test may now be the result of differential attrition.

The ninth threat to internal validity is actually a set of threats. This set is called **additive and interactive effects**.

- Additive and interactive effects refers to the fact that the threats to validity can combine to produce a bias in the study which threatens our ability to conclude that the independent variable is the cause of differences between groups on the dependent variable. They only apply to two or multigroup designs; they do not apply to the one-group design.
- These threats occur when the different comparison groups are affected **differently (or differentially)** by one of the earlier threats to internal validity (i.e., history, maturation, testing, instrumentation, or statistical regression).
- A selection-history effect occurs when an event occurring between the pretest and posttest differentially affects the different comparison groups. You can think of this as what could be called a differential history effect.
- A selection-maturation effect occurs if the groups mature at different rates. For example, first grade students may tend to naturally change in reading ability during the school year more than third grade students. Hence, part of any observed differences in the reading ability of the two groups at the posttest may

be due to maturation. You can think of this a what could be called a differential maturation effect.

- You now should be able to construct similar examples demonstrating the following:
- Selection-testing effect (where testing affects the groups differently)
- Selection-instrumentation effect (where instrumentation occurs differentially)
- Selection-regression artifacts effect (where regression to the mean occurs differentially).
- Remember that the key for the selection-effects is that the groups must be affected differently by the particular threat to internal validity.

External Validity

External validity has to do with the degree to which the results of a study can be generalized to and across populations of persons, settings, times, outcomes, and treatment variations.

- A good synonym for external validity is ***generalizing validity*** because it always has to do with how well you can generalize research results.
- The major types of external validity are population validity, ecological validity, temporal validity, temporal validity, treatment variation validity, and outcome validity. I will discuss each of these now...

Population Validity

The first type of external validity is called population validity.

- Population validity is the ability to generalize the study results to individuals who were not included in the study.
- The issues are how well you can generalize your sample results **to** a population, and how well you can generalize your sample results **across** the different kinds if people in the larger population.
- Generalizing from a sample to a population can be provided through random selection techniques (i.e., a good sample lets you generalize to a population, as you learned in the earlier chapter on sampling).
- Generalizing **across** populations is present when the result (e.g., the effectiveness of a particular teaching technique) works **across** many different kinds of people (it works for many sub populations). This is the issue of **"how widely does the finding apply?"** If the finding applied to every single individual in the population then it would have full population validity. Research results that apply broadly are welcome to practitioners because it makes their jobs easier.
- Both of these two kinds of population validity are important; however, some methodologists (such as Cook and Campbell) are more concerned about generalizing across populations. That is, they want to know how widely a finding applies.

Ecological Validity

Ecological validity is present to the degree that a result generalizes across different **settings.**

- For example, let's say that you find that a new teaching technique works in urban schools. You might also want to know if the same technique works in rural schools and suburban schools. That is, you would want to know if the technique works across different **settings**.
- Reactivity is a threat to ecological validity. Reactivity is defined as an alteration in performance that occurs as a result of being aware of participating in a study. In other words, reactivity occurs sometimes because research study participants might change their performance because they know they are being observed.
- Reactivity is a problem of ecological validity because the results might only generalize to other people who are also being observed.
- A good metaphor for reactivity comes from television. Once you know that the camera is turned on to YOU, you might shift into your "television" behavior. This can also happen in research studies with human participants who know that they are being observed.
- Another threat to ecological validity (not mentioned in the chapter) is called experimenter effects. This threat occurs when participants alter their performance because of some unintentional behavior or characteristics of the researcher. Researchers should be aware of this problem and do their best to prevent it from happening.

Temporal Validity

Temporal validity is the extent to which the study results can be generalized across time.

- For example, assume you find that a certain discipline technique works well with many different kinds of children and in many different settings. After many years, you might note that it is not working any more; You will need to conduct additional research to make sure that the technique is robust over time, and if not to figure out why and to find out what works better. Likewise, findings from far in the past often need to be replicated to make sure that they still work.

Treatment Variation Validity

Treatment variation validity is the degree to which one can generalize the results of the study across variations of the treatment.

- For example, if the treatment is varied a little, will the results be similar?
- One reason this is important is because when an intervention is administered by practitioners in the field, it is unlikely that the intervention will be administered exactly as it was by the original researchers.
- This is, by the way, one reason that interventions that have been shown to work end up failing when they are broadly applied in the field.

Outcome Validity

Outcome validity is the degree to which one can generalize the results of a study across different but related dependent variables.

- For example, if a study shows a positive effect on self-esteem, will it also show a positive effect on the related construct of self-efficacy?

- A good way to understand the outcome validity of your research study is to include several outcome measures so that you can get a more complete picture of the overall effect of the treatment or intervention.

Here is a brief summary of external validity:

- Population validity = *generalizing* to and across populations.
- Ecological validity = *generalizing* across settings.
- Temporal validity = *generalizing* across time.
- Treatment variation validity = *generalizing* across variations of the treatment.
- Outcome validity = *generalizing* across related dependent variables.

As you can see, all of the forms of external validity concern the degree to which you can make generalizations.

Construct Representation

Educational researchers must measure or represent many different constructs (e.g., intelligence, ADHD, types of on-line instruction, academic achievement).

- The problem is that, usually, there is no single behavior or operation available that can provide a complete and perfect representation of the construct.
- The researcher should always clearly specify (in the research report) the way the construct was represented so that a reader of the report can **understand** what was done and be able to **evaluate** the quality of the measure(s).
- Operationalism refers to the process of representing a construct by a specific set of operations or measures.
- For example, you might choose to represent (or "operationalize") the construct of self-esteem by using the ten item Rosenberg Self-Esteem Scale shown on page 165, and shown here for your convenience.

FIGURE 6.1 The Rosenberg Self-Esteem Scale

Circle one response for each of the following ten items.

	Strongly Agree	Agree	Disagree	Strongly Disagree
1. I feel that I am a person of worth, at least on an equal basis with others.	1	2	3	4
2. I feel that I have a number of good qualities.	1	2	3	4
*3. All in all, I am inclined to feel that I am a failure.	1	2	3	4
4. I am able to do things as well as most other people.	1	2	3	4
*5. I feel I do not have much to be proud of.	1	2	3	4
6. I take a positive attitude toward myself.	1	2	3	4
7. On the whole, I am satisfied with myself.	1	2	3	4
*8. I wish I could have more respect for myself.	1	2	3	4
*9. I certainly feel useless at times.	1	2	3	4
*10. At times I think I am no good at all.	1	2	3	4

*Items marked with an asterisk have reversed wording. The numbers on items with reversed wording should be reversed before summing the responses for the ten items. For example, on item 3, "strongly agree" becomes 4, "agree" becomes 3, "disagree" becomes 2, and "strongly disagree" becomes 1.

Source: Morris Rosenberg's "Self-Esteem Scale" from pp. 325–327 of *Society and Adolescent Self-Image* © 1989 by Morris Rosenberg, Wesleyan University Press.

- Why do you think Rosenberg used 10 items to represent self-esteem? The reason is because it would be very hard to tap into this construct with a single item.
- Rosenberg used what is called **multiple operationalism** (i.e., the use of several measures to represent a construct).
- Think about it like this: Would you want to use a single item to measure intelligence (e.g., how do you spell the word "restaurant")? No! You might even decide to use more than one test of intelligence to tap into the different dimensions of intelligence.

- Whenever you read a research report, be sure to check out how they represent their constructs. Then you can evaluate the quality of their representations or "operationalizations."

Research Validity in Qualitative Research

Now we shift our attention to **qualitative** research! If you need a review of qualitative research, review pages 45-48 in Chapter 2 for a quick overview. Also look at the qualitative research article in Appendix B titled "*You Don't Have to Be Sighted to Be a Scientist, Do You? Issues and Outcomes in Science Education*."

- One potential threat to watch out for is **researcher bias** (i.e., searching out and finding or confirming only what you want or expect to find).
- Two strategies for reducing researcher bias are reflexivity (constantly thinking about your potential biases and how you can minimize their effects) and negative-case sampling (attempting to locate and examine cases that disconfirm your expectations).

Now I will briefly discuss the major types of validity in qualitative research, and I will list some very important and effective strategies that can be used to help you obtain high qualitative research validity or trustworthiness.

Descriptive validity

Descriptive validity is present to the degree that the account reported by the researcher is accurate and factual.

- One very useful strategy for obtaining descriptive validity is investigator triangulation (i.e., the use of multiple investigators to collect and interpret the data).
- When you have agreement among the investigators about the descriptive details of the account, readers can place more faith in that account.

Interpretive validity

Interpretive validity is present to the degree that the researcher accurately portrays the *meanings given by the participants* to what is being studied.

- Your goal here is to "get into the heads" of your participants and accurately document *their viewpoints and meanings*.
- One useful strategy for obtaining interpretive validity is by obtaining participant feedback or "member checking" (i.e., discussing your findings with your participants to see if they agree and making modifications so that you represent their meanings and ways of thinking).
- Another useful strategy is to use of low-inference descriptors in your report (i.e., description phrased very close to the participants' accounts and the researcher's field notes).

Theoretical validity

Theoretical validity is present to the degree that a theoretical explanation provided by the researcher fits the data.

- I listed four helpful strategies for this type of validity.
- The first strategy is extended fieldwork (collecting data in the field over an extended period of time).
- The second is theory triangulation (using multiple theories and perspectives to help you interpret the data).
- The third is pattern matching (making unique or complex predictions and seeing if they occur; this is, did the fingerprint that you predicted actually occur?).
- The fourth strategy is peer review (discussing your interpretations and conclusions with your peers or colleagues who are not as deep into the study as you are).

Internal validity

Internal validity is the same as it was for quantitative research. It is the degree to which a researcher is justified in concluding that an observed relationship is causal. It also refers to whether you can conclude that one event caused another event. The issue of causal validity is important if the qualitative researcher is interested in making any tentative statements about cause and effect.

- I have listed three strategies to use if you are interested in cause and effect in qualitative research.
- The first strategy is called researcher-as-detective (carefully thinking about cause and effect and examining each possible "clue" and then drawing a conclusion).
- The second is called methods triangulation (using multiple methods, such as interviews, questionnaires, and observations in investigating an issue)
- The third strategy is called data triangulation (using multiple data sources, such as interviews with different types of people or using observations in different settings). You do not want to limit yourself to a single data source.

External validity

External validity is pretty much the same as it was for quantitative research. That is, it is still the degree to which you can *generalize* your results to other people, settings, and times.

- Note that generalizing has traditionally **not** a priority of qualitative researchers. However, in many research areas today, it is becoming an important goal.
- One form of generalizing in qualitative research is called naturalistic generalization (generalizing based on similarity).
- When you make a naturalistic generalization, you look at your students or clients and generalize to the degree that they are similar to the students or clients in the qualitative research study you are reading. In other words, the reader of the report is making the generalizations rather than the researchers who produced the report.
- Qualitative researchers should provide the details necessary so that readers will be in the position to make naturalistic generalizations.
- Another way to generalize qualitative research findings is through replication. This is where you are able to generalize when a research result has been shown with different sets of people, at different times, and in different settings.
- Yet another style of generalizing is theoretical generalizations (generalizing the theory that is based on a qualitative study, such as a grounded theory research

study. Even if the particulars do not generalize, the main ideas and the process observed might generalize.

Here is a summary of the strategies used in qualitative research. (Note: they are also used in mixed research and can be used creatively even in quantitative research.)

TABLE 8.2 Strategies Used to Promote Qualitative Research Validity

Strategy	Description
Researcher-as-detective	A metaphor characterizing the qualitative researcher as he or she searches for evidence about causes and effects. The researcher develops an understanding of the data through careful consideration of potential causes and effects and by systematically eliminating rival explanations or hypotheses until the final case is made beyond a reasonable doubt. The detective can utilize any of the strategies listed here.
Extended fieldwork	To provide for both discovery and validation researchers should collect data in the field over an extended time period.
Low-inference descriptors	The use of description phrased very close to the participants' accounts and researchers' field notes. Verbatims (i.e., direct quotations) are a commonly used type of low-inference descriptors.
Triangulation	Cross-checking information and conclusions through the use of multiple procedures or sources. When the different procedures or sources are in agreement you have corroboration.
Data triangulation	The use of multiple data sources to help understand a phenomenon.
Methods triangulation	The use of multiple research methods to study a phenomenon.
Investigator triangulation	The use of multiple investigators (i.e., multiple researchers) in collecting, analyzing, and interpreting the data.
Theory triangulation	The use of multiple theories and perspectives to help interpret and explain the data.
Participant feedback	The feedback and discussion of the researcher's interpretations and conclusions with the actual participants and other members of the participant community for verification and insight.
Peer review	Discussion of the researcher's interpretations and conclusions with other people. This includes discussion with a disinterested peer, (e.g., with another researcher not directly involved). This peer should be skeptical and play the devil's advocate, challenging the researcher to provide solid evidence for any interpretations or conclusions. Discussion with peers who are familiar with the research can also help provide useful challenges and insights.
External audit	Using outside experts to assess the study quality.
Negative-case sampling	Locating and examining cases that disconfirm the researcher's expectations and tentative explanation.
Reflexivity	Involves self-awareness and critical self-reflection by the researcher on his or her potential biases and predispositions as these may affect the research process and conclusions.
Pattern matching	Predicting a series of results that form a distinctive pattern and then determining the degree to which the actual results fit the predicted pattern or "fingerprint."

The bottom line of this chapter is this: You should always try to evaluate the research validity of empirical studies before trusting their conclusions. And, if you are conducting

research you must use validity strategies if your research is going to be trustworthy and defensible.