

PSY 102 – Understanding and Thinking Critically About Psychological Research

Thinking critically about research means knowing the right questions to ask to assess the validity or accuracy of a knowledge claim. Many times there will not be anyone present (or the necessary information available) to adequately answer the questions you raise. However, stopping to ask the questions will lead you to be appropriately skeptical or cautious of particular knowledge claims (which are typically made with some sense of authority) until they can be substantiated. Note that thinking critically is *not* necessarily the same as criticizing someone or someone's research. Thinking critically about research leads to appreciation of the strengths as well as the weaknesses of a particular study. Instead of criticism, critical thinking is based on being curious, asking relevant questions, and challenging assumptions.

Critical thinking also includes taking a logical approach to understanding the topic, being careful to consider how our emotions, beliefs, and biases, as well as the emotions, beliefs, and biases of others, can influence how research is conducted, data are gathered, and results are interpreted. Thinking critically about research involves trying to separate facts from opinions, and trying to examine available evidence as objectively as possible. Using sound logic to draw conclusions from the evidence requires the ability to consider alternative explanations for the research findings. Might there be another explanation for the data other than the one that is being provided or the one that seems "obvious" to us? Typically there are *several* alternative explanations or interpretations of any behavior or research finding, and being a good critical thinker means being willing to accept that ultimate answers to questions about human behavior are not liable to be simple or clear-cut. Oftentimes we are left with more questions than when we began studying a particular topic.

Why do Psychologists Perform Research?

The first question to ask when faced with a research study is, "What are the goals of the research?" Of course for each study there may be several answers to this question. Sometimes the goal is simply to discover and describe what exists. Such *descriptive* studies are conducted to determine the frequency of particular behaviors or experiences, or to document public opinion about some issue. If the phenomenon being studied is rare, researchers may have access only to an individual or small group of people. This form of descriptive research is typically referred to as *case studies*. Sometimes the goal of research is to discover relationships among variables. For example, is there a relationship between age and attitudes toward the homeless? A *correlational* study would answer this question. Sometimes the goal of research is to determine what causes or influences a particular behavior. Such *experimental* studies involve manipulating conditions to observe what effect, if any, the manipulation has on behavior. We will examine each of these types of research, and their strengths and weaknesses, later in this reading. First, how do we determine why a particular study was conducted?

Often the reason for the research is contained in the hypothesis statement. An *hypothesis* is simply a research question or statement, such as "Do men and women differ in the views of romantic love?" or "What is the relationship between depression and feelings of loneliness?" Why is it important to consider the stated reason the research was conducted? Such statements tell us, the reader, the primary focus of the research as well as what the researchers hoped to learn from their work. This information will be helpful in evaluating whether the research design

was adequate to answer the primary research questions and whether the writers are justified in their interpretation of their findings and the implications they draw from them. In a nutshell, the research questions the writers pose tell us what we can expect to learn from reading about the research and cues us into the aspects of the research we should examine with special attention.

What if one hears about research findings on television or reads about them in a magazine, newspaper, or textbook? Here we are usually not told the reason or objective of the research. In these cases it can be just as useful to turn our evaluative question around. Instead of asking, "Why was the research conducted?," we could ask, "What kind of research would one have to conduct to accurately arrive at that conclusion?" To answer that question we need to be familiar with the various types of research designs, and the strengths and weaknesses of each.

Kinds of Research

Descriptive studies are meant to describe the state of affairs with regard to a certain behavior, experience, or attitude. Research designed to determine parents' attitudes toward their children, the percentage of people who have experienced domestic violence, or the frequency fighting among dating couples are all examples of descriptive studies. Descriptive studies cannot tell us anything about relationships among variables or what causes particular attitudes or behavior. However, research on any particular topic in psychology generally began with descriptive studies that documented that there was something that needed further study. **Case studies** are one form of descriptive research. Sometimes a particular experience or disorder is so rare that it is difficult or impossible to find a group of such people to study. In cases studies, and individual or small set of people are studied in great depth. Again, because case studies are simply descriptive, we cannot determine relationships among variables or what caused something else. Also, because case studies involve an individual or a small set of people, the extent to which the findings generalize to other people is unknown.

Many times researchers are interested in trying to determine what causes something else. We say research is **experimental** when the researchers manipulate some variable(s) and examine the effect of that manipulation. For example, researchers interested in the effects of exposure to violent films on attitudes toward violence might measure such attitudes in two groups: the first group was not exposed to any films, or perhaps only to films with nonviolent content, whereas the second group was shown a set amount of violent film. If participants in the study were randomly assigned to either of the two groups, then the researchers assume that they are similar in all aspects except exposure to violence in films. So, if the two groups subsequently differ in their reported attitudes toward violence, the researchers conclude that exposure to the films affected the participants' attitudes.

Notice that a hallmark of experimental research is the control group (the group who is not exposed to the experimental manipulation). It is important to have a control group for comparison. If all the researchers have is an experimental group, and there does seem to be some change from before the exposure to afterwards, how would they know whether the change was due to the experimental manipulation? It might be that the participants would have changed in that way anyway. A control group allows for examination of that possibility.

Ideally a control group would be exposed to all of the same experiences as the experimental group, except for whatever is the actual experimental manipulation. To get back to the example of exposure to film with violent content, suppose the control group was not shown any films at all, but instead were simply surveyed as to their attitudes toward violence (and their scores on this measure were to be compared to scores on the same measure by participants who saw the violent films). If there was a difference between the experimental and control groups, would it be because of having seen violent films, or simply watching films period, or being in an experiment where the participants expected that something would happen? There would be no way of knowing. So, in this case, the control group should also watch films (for the same length of time as the experimental group). That way both groups are equal in the sense that they watched films and realized that they were in a study.

This last issue, about knowing that one is in an experiment, has to do with *expectancy effects*. When people take part in research (especially an experiment where they are being exposed to some manipulation), they tend to expect that something will happen to them. Or they may try to "figure out" what it is that the researchers expect. Research participants may not even do this very consciously, but the risk is that they will behave a certain way, or report certain things, because they believe that is what is expected of them. For this reason, it is often necessary to disguise the "real" meaning of an experiment, so that research participants do not act accordingly. When these participant expectation effects involve the belief that the manipulation will improve their condition in some way, researchers refer to them as *placebo effects*. For example, if research participants take part in a study where it seems that the purpose is to affect student study skills, participants will tend to believe that their study skills have been improved as a result of the experimental manipulation, even if there is no actual improvement in the study skills. It is not that the participants are lying or trying to deceive the researchers, but they tend to believe they have improved because, after all, they went through some intervention that seems like it should have improved their study skills.

Expectancy effects and placebo effects are potential weaknesses of experiments. Another potential weakness of experimental research is that often the experimental manipulation does not match well with reality. In the current example of exposure to violent films, it is unlikely that the amount of exposure to violent film in the experiment would be anywhere close to the amount experienced by people over the course of their lives. Because the exposure to violent film in the experiment is limited, only one or a few particular film clips will be used. Are these representative of all violent films that are available to consumers? Also, if the participants are randomly assigned to the two experimental conditions, then many of those who are exposed to the violent films are not people who would expose themselves outside of the research setting. The way these people are affected by exposure to violent film may not mirror how people who seek out violent films in the "real world" would be.

Sometimes researchers are not able to perform random assignment of research participants, and must make do with the groups available. For example, suppose researchers were interested in evaluating the effect of a therapy program on depression. In an experiment, people seeking treatment for depression would be randomly assigned to a treatment group or a control group that did not receive treatment. However, suppose that, as potential research participants telephoned a mental health clinic seeking help, the researchers were not allowed to deny treatment so that they could form a comparison group. Luckily for the researchers, at some

point the clinic received more patients than can be treated at one time, so people calling after that were put on a waiting list.

Now the researchers can examine the degree of improvement in depression between the patients receiving treatment and those waiting to receive treatment for depression. Because one group is being treated differently than the other group, this type of study resembles an experiment. However, because there was no random assignment, we refer to this research design as ***quasi-experimental***. Unfortunately, without random assignment to the two groups, the researchers cannot be sure that the groups are alike in all respects other than having received treatment. There is the possibility that differences in the rate of improvement between the two groups might be due to something other than treatment (like degree of motivation--those who actually entered treatment may be more motivated than those who only got as far as calling the clinic). Still, a quasi-experiment is better than no research at all.

In contrast to the experiment, most of the research in psychology and other behavioral sciences is ***correlational***. That is, variables are not manipulated, but rather the researchers measure at least two variables and examines relationships between them. The resulting correlation coefficient (abbreviated *r*) can range from -1.00 to 1.00. The presence or absence of a negative sign simply indicates the *direction* of the relationship between the two variables and has *nothing* to do with the strength of the relationship. A negative sign indicates that as the value of one variable increases the value of the other variable decreases; whereas the absence of a negative sign indicates that as the value of one variable increases the other increases as well. So, a correlation of -.42 actually indicates a stronger relationship between two variables than does a correlation of .27.

Returning to our earlier example, if the focus research is possible relationships between exposure to violent film and attitudes toward violence, the researchers might ask people to complete a questionnaire containing measures of exposure to violent films (e.g., "How often during the past six months did you watch videos, movies, or television shows that contained explicit violence?") as well as attitudes toward violence. To the extent that scores on the two measures are related, the researchers conclude that violent films *may* affect attitudes.

When interpreting results of correlational research, conclusions as to cause-and-effect are speculative at best. All the researchers can say for sure is that the two variables appear to be related. However, because the researchers did *not* manipulate either variable, it is unknown which variable affects the other, or whether both are due to some other (third) variable or set of variables. In our current example, does exposure to violent films lead people to hold different attitudes toward violence, or are people who hold particular attitudes toward violence most likely to then seek out violent films? Alternatively, perhaps people who are most religious happen to hold particular attitudes toward violence *and* happen to be least likely to seek out violence films, so attitudes toward violence and exposure to violent films may not affect one another, but may be correlated with one another because both vary according to the religiosity of the individual.

How do Researchers Obtain Participants for Their Research?

Researchers start with a *population* of interest, whether it be all people, or adolescent boys, or African-American lesbian women. Because it is typically impossible, or at least not feasible, to study every member of the population of interest, researchers must rely on studying a selected group, or *sample*, of those individuals. Results are *generalizable* to the extent that the findings of the study (which are based on the sample) mirror what one would find in the larger population of interest. Ideally, a research sample would be perfectly *representative* of the population of interest (all members of the population would have an equal chance of being included in the research). In reality, this is impossible to achieve as there are always some potential participants who refuse to take part in the research—this is just one of the things researchers live with, regardless of their field of study.

The larger concern is the extent to which a particular sample deviates from the population of interest. If the people in the sample differ in important ways from the larger population from which they were drawn, we should question the extent to which the findings apply to the people in the population who were not part of the research sample. We might refer to such a sample as *biased* in the sense that it is not truly representative. Notice that we aren't saying the individuals in the sample are biased toward the topic that is being studied, but rather that the sample as a whole is biased toward including more of a particular type of person than exists in the population.

Volunteer Bias

Research results are often presented in such a way as to imply that they accurately describe *people in general*, or at least all people in the population of interest. However, not everyone from the population of interest is given the opportunity to participate in research and not everyone who is given the opportunity actually agrees to do so. Because people are free to decline an invitation to participate in research, some people will choose that option, perhaps because they do not have the time or interest to participate. This is true about research in general, but we can imagine how the issue may be most relevant when the research is on a sensitive topic or involves a great deal of time or effort on the part of research participants. So, perhaps it is not surprising that even in the most extensive and well-conducted national surveys, where great care is taken to select a nationally-representative sample, only about 70-80% of those people initially selected to participate actually do so.

Are there differences between those individuals who agree to participate in research and those who do not? Typically yes, although how volunteers and nonvolunteers differ is not always known. It does seem that the more sensitive the topic, or the more demanding the requirements of participation, the more likely a sample is going to be very different from people in the general population. In other words, the more unusual or time-consuming the study, the more likely the sample will be biased (that is, not representative of the population).

College Student Participants

It is important to note that a large proportion of published research reports in psychology are based on college student participants. Students are often a captive audience, and since most

psychology researchers are faculty members, it makes sense that they would turn to college students as research participants. One may question how representative college students are of the general population. First, many college students are only 18 to 22 years old, which limits the range of experience they are likely to have had compared to older adults. Perhaps even more troubling is the possibility that samples of college students employed in research differ even from those college students who do not volunteer for such research. A common method of recruiting potential college student participants involves use of a *subject pool* generated because some college courses require participation in research to pass the course or offer extra credit for research participation. These subject pools often are part of introductory social science courses, such as Introduction to Psychology. How does use of such subject pools result in a potentially biased sample of college students?

Courses such as Introductory Psychology may attract certain types of college students and not others (so not all college students at a particular campus have a chance to participate in the research). Also, many times the sign-up procedures involved in recruiting individual participants from the pool include at least a brief description of the study and what participation entails. The rationale for such descriptions is that students have a right to know what to expect when they are selecting studies in which to potentially participate (the principle of *informed consent*). The problem is that such descriptions attract certain students and repel others, leading to a unique sample of college students for particular types of studies. That is, certain types of research will appeal to some college students and not others.

How do Researchers Define and Measure Variables?

Psychologists are interested in understanding behavior and certain concepts (such as self-esteem or personality). However, when they perform research it is possible to consider all possible ways to define or measure a particular behavior or concept. So, each study represents a decision that was made as to how variables would be defined and measured. The definitions and measures that are chosen can have a large impact on the results.

For example, suppose researchers are interested in the concept altruism. There are several aspects of altruism, including willingness to help others, to place others' needs before one's own, and so forth. So, for one particular study, researchers are probably going to have to get a bit more specific. Suppose that one set of researchers decided to focus on "helping behavior," and even more specifically, helping a stranger in need. Even here there are several different ways someone could define "helping a stranger in need" or could try to measure this behavior. These researchers decide to have an actor "fake" a seizure in a shopping mall, and researchers will measure helping responses as a function of shopper's gender, age, and how crowded it is in the mall. In this way the researchers hope to better understand how gender, age, and size of audience may influence helping behavior. There are still measurement questions that need to be answered. What counts as "helping behavior"? How will the level of crowdedness be measured? Will the shoppers be asked their age or will it be estimated based on appearance?

The answers to each of these questions for the *operational definitions* of the variables in this study. Researchers start with concepts, such as "helping behavior," and "age," and "crowdedness," but must always *operationally define* how each concept will be measured. In this example, suppose another set of researchers was also interested in how gender, age, and

size of audience affected helping behavior. However, they decide to administer questionnaires to people waiting in airports. The questionnaire asks for the respondent's age and gender, and then describes several scenarios. For each scenario the respondent is asked to rate how likely it is that he or she would help the stranger in need who is described in the scenario. In this study, the operational definition of "helping behavior" is the ratings of likelihood of helping in each scenario--quite different than the operational definition of helping behavior in the first study.

Now suppose that you read that some researchers (the first set) did not find a gender difference in helping behavior, but that other researchers (the second set) had. If you did not know that the two sets of researchers had used different operational definitions of "helping behavior," you would not have any way of making sense out of the contradictory findings. In this example, which operational definition do you think most closely matches helping behavior in "the real world"?

Frequently, psychological research relies on peoples' self-reports. We should always be cautious in taking such self-reports at face value. First, people may not have good insight into what they are being asked to report. For example, if you were asked why you chose Columbia College (or your mate, or your major), you may come up with an answer, but how completely would that verbal answer capture all of the complexity that went into your choice? Often people are asked to report things that happened long ago (during childhood?) or to report how frequently something has happened. Later in the semester we will study memory, but at this point we should recognize the limits of human memory to accurately recall certain things. Then there are issues of honesty. People tend to be motivated to present themselves in a positive way. We do not like to think negatively of ourselves, nor do we want others to think negatively of us. So, we may be motivated to remember things a certain way, or to distort our responses to put the best possible light on them. We can expect people to "over-report" socially desirable behavior and attitudes, because to report otherwise would mean admitting something negative.