External Validity Cards

(adapted from study.com)

(1) External validity - definition and goals

External validity is the extent to which results of a study can be generalized to the world at large.

The goal of research is to

a. to make inferences about the way things work in the real world based on the results of a study,

b. **generalize** to the population as a whole, based on an experiment done on a small sample of the population, or

c. **generalize** from a task done in a lab to a real-world setting, like an office or a school

Low external validity means that the research done doesn't tell us anything about the world outside of the study. That's a very limited viewpoint!

(2) External validity vs. Internal validity

So why not do research in the field with the actual target users and have a strong relationship to the real world and **generalizing** results?

a. No access to real users

b. Because, doing research in the lab allows for a lot of **control** over the variables and a strong argument that only your independent variable matters i.e., **internal validity**.

There are always payoffs in research, and every study has to find the right balance between internal and external validity, or control and generalization. What that balance is depends on the goals of the researcher and the study itself.

(3) External validity requirements

1. Replication is when a study can be done again and the same general results are found. If a study is highly replicable, that means that it can be done over and over and the same result will be found each time.

Replication is important for external validity because if a study can't be replicated, how can we know that the results are true?

And if researchers tweak an experiment (perhaps by using different types of subjects, varying the task, or doing it in a non-lab setting), and then they get the same result, then the external validity of the study is even stronger. After all, if the results hold true for a variety of people and settings, there's a good chance that they can be generalized to the larger population.

2. Internal validity: extent to which a researcher can say that only the independent variable is causing the dependent variable.

(4) External validity - Types of generalization

1. **Across subjects:** get a sample that represents the population, and be able to get the same results with a different group of participants.
2. **Producers that represent constructs:** how do you measure something that can't be observed? In psychological research, **constructs** are measured by choosing an observable behavior or behaviors to represent the construct. Lets say you want to measure engagement in a classroom. You estimate that you can measure this construct, engagement, by counting the number of times students raise their hands. Though this is not actually engagement, it is related to and perhaps indicative of engagement. Generalizing from procedures to constructs involves measuring the abstract **construct** in your study correctly.
3. **Generalization beyond one task and the lab:** when you can say that your results would be the same in a different, real-world setting.

(5) External validity - Threats

1. **Sample characteristics** is when the subjects chosen for the experiment interfere with the results. The smaller the **sample size** of an experiment (that is, the fewer the people studied), the less likely it is that they represent the population as a whole. Besides sample size, another sample characteristic that can affect external validity is **selection bias**, which occurs when only certain people are chosen for an experiment.

(6) External validity - Threats

2. **Stimulus characteristics:** when the task the users do in an experiment doesn't represent a real world task. For example, if you assume that puzzle solving and problem solving require the same human skill, and then you use puzzle solving speed to infer about people's problem solving capabilities your external validity may be very low, if that is untrue.

(7) External validity - Threats

3. **Multiple-treatment interference:** Essentially, the question to ask when looking for multiple-treatment interference is: Did one of the treatments affect the other? If the answer is yes, then that could be a problem for the results.

(8) External validity - Threats

4. **Novelty effects:** occur when the results of a study are due to the novelty (or newness) of a treatment. Essentially, in the case of novelty effects, anything different makes a difference. Novelty effects pose a threat to external validity because they make it difficult to know if the results of the study are due to a treatment that works or due to the novelty of a treatment. In other words, does a new app actually teach math, or do the users learn math faster just because the app is different?

(9) External validity - Threats

5. **Test sensitization:** For example, let's say that you wanted to know how much math the students knew before they use the new learning app. You might give them a math test at the beginning of the study to see what they know. But what if the test you give the students has a lot of questions about fractions on it? They might realize that they're supposed to be learning fractions, so when they play with the app, they pay attention to the fractions part of it. The app itself might not work as well as it does when the students are given a test.

   Test sensitization is a problem because there's no way that you can separate out the effects of the treatment alone and the effects of the treatment and the test.

(10) External validity - Threats

6. **Measurement timing:** When the effects of a treatment only last a short time. Measurement timing is a problem because if your app only improves math for a short period of time, it might not be worth it to use in classrooms.

(11) External validity - Threats

7. **Reactive arrangements** are when subjects change their behavior because they are participating in an experiment. In other words, they don't act the way they normally would. Another scenario where reactive arrangements come into play is when behaviors are being observed.

   You might be thinking that researchers could get around reactive arrangements by just not telling the subjects that they are in an experiment. The problem with this is that it violates the ethics of research. All research studies require informed consent of the participants. Alternatively, you can uses **nonreactive measures,** observations done when subjects don't know that they are being observed.