

1.12: Assessing the validity of a study

More than any other thing, a scientist wants their research to be “valid”. The conceptual idea behind **validity** is very simple: can you trust the results of your study? If not, the study is invalid. However, while it’s easy to state, in practice it’s much harder to check validity than it is to check reliability. And in all honesty, there’s no precise, clearly agreed upon notion of what validity actually is. In fact, there’s lots of different kinds of validity, each of which raises it’s own issues, and not all forms of validity are relevant to all studies. I’m going to talk about five different types:

- Internal validity
- External validity
- Construct validity
- Face validity
- Ecological validity

To give you a quick guide as to what matters here...(1) Internal and external validity are the most important, since they tie directly to the fundamental question of whether your study really works. (2) Construct validity asks whether you’re measuring what you think you are. (3) Face validity isn’t terribly important except insofar as you care about “appearances”. (4) Ecological validity is a special case of face validity that corresponds to a kind of appearance that you might care about a lot.

Internal validity

Internal validity refers to the extent to which you are able draw the correct conclusions about the causal relationships between variables. It’s called “internal” because it refers to the relationships between things “inside” the study. Let’s illustrate the concept with a simple example. Suppose you’re interested in finding out whether a university education makes you write better. To do so, you get a group of first year students, ask them to write a 1000 word essay, and count the number of spelling and grammatical errors they make. Then you find some third-year students, who obviously have had more of a university education than the first-years, and repeat the exercise. And let’s suppose it turns out that the third-year students produce fewer errors. And so you conclude that a university education improves writing skills. Right? Except... the big problem that you have with this experiment is that the third-year students are older, and they’ve had more experience with writing things. So it’s hard to know for sure what the causal relationship is: Do older people write better? Or people who have had more writing experience? Or people who have had more education? Which of the above is the true *cause* of the superior performance of the third-years? Age? Experience? Education? You can’t tell. This is an example of a failure of internal validity, because your study doesn’t properly tease apart the *causal* relationships between the different variables.

External validity

External validity relates to the **generalizability** of your findings. That is, to what extent do you expect to see the same pattern of results in “real life” as you saw in your study. To put it a bit more precisely, any study that you do in psychology will involve a fairly specific set of questions or tasks, will occur in a specific environment, and will involve participants that are drawn from a particular subgroup. So, if it turns out that the results don’t actually generalize to people and situations beyond the ones that you studied, then what you’ve got is a lack of external validity.

The classic example of this issue is the fact that a very large proportion of studies in psychology will use undergraduate psychology students as the participants. Obviously, however, the researchers don’t care *only* about psychology students; they care about people in general. Given that, a study that uses only psych students as participants always carries a risk of lacking external validity. That is, if there’s something “special” about psychology students that makes them different to the general populace in some *relevant* respect, then we may start worrying about a lack of external validity.

That said, it is absolutely critical to realize that a study that uses only psychology students does not necessarily have a problem with external validity. I’ll talk about this again later, but it’s such a common mistake that I’m going to mention it here. The external validity is threatened by the choice of population if (a) the population from which you sample your participants is very narrow (e.g., psych students), and (b) the narrow population that you sampled from is systematically different from the general population, *in some respect that is relevant to the psychological phenomenon that you intend to study*. The italicized part is the bit that lots of people forget: it is true that psychology undergraduates differ from the general population in lots of ways, and so a study that uses only psych students *may* have problems with external validity. However, if those differences aren’t very relevant to the

phenomenon that you're studying, then there's nothing to worry about. To make this a bit more concrete, here's two extreme examples:

- You want to measure “attitudes of the general public towards psychotherapy”, but all of your participants are psychology students. This study would almost certainly have a problem with external validity.
- You want to measure the effectiveness of a visual illusion, and your participants are all psychology students. This study is very unlikely to have a problem with external validity

Having just spent the last couple of paragraphs focusing on the choice of participants (since that's the big issue that everyone tends to worry most about), it's worth remembering that external validity is a broader concept. The following are also examples of things that might pose a threat to external validity, depending on what kind of study you're doing:

- People might answer a “psychology questionnaire” in a manner that doesn't reflect what they would do in real life.
- Your lab experiment on (say) “human learning” has a different structure to the learning problems people face in real life.

Construct validity

Construct validity is basically a question of whether you're measuring what you want to be measuring. A measurement has good construct validity if it is actually measuring the correct theoretical construct, and bad construct validity if it doesn't. To give very simple (if ridiculous) example, suppose I'm trying to investigate the rates with which university students cheat on their exams. And the way I attempt to measure it is by asking the cheating students to stand up in the lecture theatre so that I can count them. When I do this with a class of 300 students, 0 people claim to be cheaters. So I therefore conclude that the proportion of cheaters in my class is 0%. Clearly this is a bit ridiculous. But the point here is not that this is a very deep methodological example, but rather to explain what construct validity is. The problem with my measure is that while I'm *trying* to measure “the proportion of people who cheat” what I'm actually measuring is “the proportion of people stupid enough to own up to cheating, or bloody minded enough to pretend that they do”. Obviously, these aren't the same thing! So my study has gone wrong, because my measurement has very poor construct validity.

Face validity

Face validity simply refers to whether or not a measure “looks like” it's doing what it's supposed to, nothing more. If I design a test of intelligence, and people look at it and they say “no, that test doesn't measure intelligence”, then the measure lacks face validity. It's as simple as that. Obviously, face validity isn't very important from a pure scientific perspective. After all, what we care about is whether or not the measure *actually* does what it's supposed to do, not whether it *looks like* it does what it's supposed to do. As a consequence, we generally don't care very much about face validity. That said, the concept of face validity serves three useful pragmatic purposes:

- Sometimes, an experienced scientist will have a “hunch” that a particular measure won't work. While these sorts of hunches have no strict evidentiary value, it's often worth paying attention to them. Because often times people have knowledge that they can't quite verbalize, so there might be something to worry about even if you can't quite say why. In other words, when someone you trust criticizes the face validity of your study, it's worth taking the time to think more carefully about your design to see if you can think of reasons why it might go awry. Mind you, if you don't find any reason for concern, then you should probably not worry: after all, face validity really doesn't matter much.
- Often (very often), completely uninformed people will also have a “hunch” that your research is crap. And they'll criticize it on the internet or something. On close inspection, you'll often notice that these criticisms are actually focused entirely on how the study “looks”, but not on anything deeper. The concept of face validity is useful for gently explaining to people that they need to substantiate their arguments further.
- Expanding on the last point, if the beliefs of untrained people are critical (e.g., this is often the case for applied research where you actually want to convince policy makers of something or other) then you *have* to care about face validity. Simply because – whether you like it or not – a lot of people will use face validity as a proxy for real validity. If you want the government to change a law on scientific, psychological grounds, then it won't matter how good your studies “really” are. If they lack face validity, you'll find that politicians ignore you. Of course, it's somewhat unfair that policy often depends more on appearance than fact, but that's how things go.

Ecological validity

Ecological validity is a different notion of validity, which is similar to external validity, but less important. The idea is that, in order to be ecologically valid, the entire set up of the study should closely approximate the real world scenario that is being investigated. In a sense, ecological validity is a kind of face validity – it relates mostly to whether the study “looks” right, but with a bit more rigour to it. To be ecologically valid, the study has to look right in a fairly specific way. The idea behind it is the intuition that a study that is ecologically valid is more likely to be externally valid. It’s no guarantee, of course. But the nice thing about ecological validity is that it’s much easier to check whether a study is ecologically valid than it is to check whether a study is externally valid. An simple example would be eyewitness identification studies. Most of these studies tend to be done in a university setting, often with fairly simple array of faces to look at rather than a line up. The length of time between seeing the “criminal” and being asked to identify the suspect in the “line up” is usually shorter. The “crime” isn’t real, so there’s no chance that the witness being scared, and there’s no police officers present, so there’s not as much chance of feeling pressured. These things all mean that the study *definitely* lacks ecological validity. They might (but might not) mean that it also lacks external validity.

This page titled [1.12: Assessing the validity of a study](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Matthew J. C. Crump](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.