Pre-lecture Notes I.5 – Threats to Construct Validity

One of the best ways to think about each of the four kinds of validity is in terms of the issues that can cause them to be low. (This assumes that you already know the definition, of course.) In other words: feel free to *start* with the assumption that your research has a sufficient level of all four types of validity to be valuable to the field, but always ask whether this is really true before going too far. In order to do this – in order to ask whether you really do have a sufficient amount of a given type of validity – you must be aware of what can cause problems for each type of validity. These are called "threats to validity" and each of the four types of validity has its own set of major (and minor) threats.

With regard to construct validity, recall that the best definition includes two components (which it got from the older idea of content validity): the measure should be exhaustive in that it covers all aspects of the target theoretical construct, and the measure should be selective in that it only covers aspects of the target theoretical construct. These are convergent and discriminant validity, respectively. Thus, the first two major threats to construct validity are these: the measure fails to be exhaustive and/or the measure fails to be selective.

Now, convergent and discriminant validity are both assessed using correlations (which comes from another older idea known as predictive validity). While it is true that all forms of validity are "extent to which…" issues and not all-or-none, there are certain rules that we use as minimum or maximum cutoffs. To be clear: a given measure must produce certain correlations before you can safely proceed. The cutoffs that psychology uses are +.70 and ±.20 for convergent and discriminant validity, respectively. The measure must have +.70 or better correlations with all existing measures of the same theoretical construct, and cannot have a correlation above +.20 or below –.20 with any existing measure of a different theoretical construct. (Note: if you've already gotten into the habit of squaring correlations so that you're dealing with the real strength of the relationship, then the +.70 and ±.20 correspond roughly to about 50% and 5%, respectively, which might be easier to remember.)

If the measure fails a test for convergent validity, something will probably have to be added, because it is not measuring all of the construct. If the measure fails a test for discriminant validity, something will probably have to be removed, because it's measuring something that isn't part of the construct.

Of course, there are some exceptions to the above rules. These will be covered in lecture and they mostly concern the type of construct that you are trying to measure. Right now, I want to focus on a new issue that doesn't depend on what you are measuring and is always a potential serious threat to construct validity. It isn't obvious from the definition of construct validity. It is also our first specially-named threat to a type of validity:

Reactivity – *any change in the behavior of subjects due to the fact that they are being measured*

(Note that some people like to substitute "in a psychology experiment or study" for "being measured" to make it clear that reactivity includes any changes in the subject due to any aspect of the experiment or study, as opposed to just focusing on the fact that the subject is being observed or is answering questions. That's fine, too. The general idea is what matters: people don't always act naturally when psychologists are around, and this is a serious problem for us.)

There are actually many sub-types of reactivity, but we will only discuss this threat in general terms for now. It can probably best be explained by example.

Imagine a subject in a study on depression that is using the BDI score as the measure. Here are the four options for one of the 21 questions on the original version of the BDI test:

- I do not feel sad.
- I feel sad.
- I am sad all the time and I can't snap out of it.
- I am so sad or unhappy that I can't stand it.

Assume, for the moment, that the correct (i.e., honest) answer for this particular subject on this question is the third. But also imagine that the subject (for whatever reason) is physically attracted to the researcher and also (for whatever reason) believes that saying "I am sad all the time and I can't snap out of it" would make the researcher less likely to agree to go out on a date with the subject. So, the subject says "I feel sad," instead of the honest answer. This is a form of reactivity. The subject is "reacting" to being measured, which includes any possible consequences of the data being available to other people, and is not answering the question completely honestly. Because of this, the question isn't really helping to provide an accurate estimate of the subject's level of depression; instead, the question is providing information about what the subject thinks is an attractive or positive attribute of people in general. At best, the way that the subject is answering the question has lowered the discriminant validity of using the BDI to measure depression, because it now also includes what is often called "social desirability," as well. At worst, the way that the subject is answering may have greatly reduced the convergent validity, too, because now the subject is answering entirely in terms of what would be attractive to other people, instead of providing information about their own internal state of mind.

Reactivity is a big problem. In lecture we'll talk about some of the ways that we try to prevent it from happening. For now, try to generate some suggestions of your own.