Pre-lecture Notes IV.1 – Paradigm Measures

The word *paradigm* has several meanings. When it's applied to an entire sub-field of a science, it refers to the unquestioned assumptions with which everyone within the sub-field agrees. For example, there's the Information Processing Paradigm (which includes quite a bit of cognitive psychology) which involves three main assumptions: that the mind is composed of several separate sub-units (such as the visual system vs the motor system); that the mind of all "normal" people have the same set of sub-units; and that these units need time to do whatever they do. As you can see, the assumptions that underlie a given paradigm (of this sort) are relatively uncontroversial. With all that said, this doesn't have much to do with research methods. The way in which sub-field paradigms come into contact with methods is that one or more of the assumptions usually determines what measures can and should be used. In the case of the Information Processing Paradigm, the third assumption – namely, that thinking takes time – is why researchers in this sub-field often either (a) give subjects as much time as they need to do the task and measure how long it takes them to do it correctly [response time] or (b) limit the amount of time that a subject is given and measure how often they respond correctly [accuracy].

Another meaning of paradigm has a much closer link to research methods because it acts as a special kind of operational definition. In this case, the thing that is a paradigm has some amount of construct validity. An example here is the Stroop paradigm. Recall from Elementary Psych that the Stroop task is the one where the subjects are supposed to name the ink colors as fast as they can while ignoring the words that are written in the ink. Sometimes the word matches the correct response, such as the word *red* written in red ink: **red**; sometimes the word doesn't match the correct response, such as the word *blue* written in green ink: **blue**. Trials on which the word matches the color are called "compatible" or "congruent"; trials on which the two don't match are "incompatible" or "incongruent."

The way that the Stroop task can become the Stroop paradigm is as follows: if you assume that the reason why the word has an effect on response time (and accuracy) is because people can't help but read the word, then the size of the difference in performance between congruent and incongruent trials provides an estimate of how automatic reading is. If response times on congruent trials are much lower than response times on incongruent trials, then reading (in this particular situation) must be very automatic. If the difference is very small or zero, then reading isn't automatic.

Note how the key piece of data on which you focus is not performance in either the congruent or incongruent condition on its own; rather, the key piece of data is the difference between the two. This difference is referred to as the effect; in this case, the difference in response time between congruent and incongruent trials is the Stroop effect. Putting it all together: the magnitude of the Stroop effect is the operational definition of the automaticity of reading. (Alternatively, if you prefer the other order: the automaticity of reading is operationally defined as the magnitude of the Stroop effect.)

Because we're talking about an operational definition, it's clear that we're in the realm of construct validity. To the extent that the Stroop effect is an exhaustive and selective measure of the automaticity of reading, then we have high levels of both convergent and discriminant validity and all is good. To the extent that the Stroop effect doesn't cover all aspects of the automaticity of reading, then convergent validity drops; to the extent that the Stroop effect can be influenced by things other than the automaticity of reading, then discriminant validity drops. (*Try to guess which of these hasn't been checked enough!*)

But there's more to a paradigm measure (as these are often called) than just construct validity. Because paradigm measures are effects – i.e., they are differences in behavior between two conditions – there is also the question of whether there is one and only one difference between the conditions from which we get the data to calculate the effect. For example, if congruent and incongruent trials differ not only in whether the word matches the correct response, but also differ in some other (unintended) way, as well, then we'd have a confound and we wouldn't be sure if the effect that we calculated is coming from the congruence between the word and response or from the other difference. In short: a paradigm measure not only has a certain level of construct validity in terms of how the data are interpreted, but the manipulation that is used to create the conditions (to get the data to calculate the effect) has a certain level of internal validity, as well.

When the construct validity of the underlying operational definition and the internal validity of the standard manipulation have both been checked and found to be sufficient, then you say that the paradigm has been "validated." One without the other won't do. You must have both. This implies that validating a paradigm is a lot more work that validating a simple measure (which only requires construct validity) or validating a new manipulation for creating a set of conditions (which only requires internal validity). But the extra work is often worth it, because many of the more interesting constructs – including most forms of automaticity, the ability to take other peoples' perspectives, and the ability to ignore distraction and remained focused, etc – all involve paradigm measures.