The glue of (ab)normal mental life: Networks of interacting thoughts, feelings and behaviors
Cramer, A.O.J.

Citation for published version (APA):
Cramer, A. O. J. (2013). The glue of (ab)normal mental life: Networks of interacting thoughts, feelings and behaviors

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Appendix E

Validity from a network perspective

Abstract

This commentary was written in response to a target article by Newton (2012, published in the same issue of Measurement) that provided a comprehensive review of the history of validity as well as a proposal for a clearer and more precise definition of validity. In this commentary, I elaborate on how the definition of validity might change, more dramatically than proposed by Newton, when adopting a network stance towards certain psychological phenomena.

What is validity? A simple question but apparently one with many answers, as Newton highlights in his review of the history of validity. The current definition of validity, as entertained in the 1999 Standards for Educational and Psychological Testing is indeed a consensus, one between the classical notion of attributes, and measures thereof, and Cronbach and Meehl’s (1955) proposition that validity should be concerned with interpretations of test scores and not with the tests themselves. Newton is certainly right when striving for a clearer and more precise definition of validity than the Standards’ definition; and in many ways Newton succeeds in achieving this goal.

In both Newton’s proposed clarification as well as the current Standards’ definition, validity is closely associated with measurement of certain attributes. For many researchers in psychology, these terms are intimately associated with the measurement model that underlies the generic latent variable model: for example, insomnia (inso), fatigue (fati), concentration problems (conc) and depressed mood (mood) are caused by the latent attribute major depression (MD). Likewise, intelligence is hypothesized to be the common cause of people’s responses to the items of an intelligence test. As such, when assessing the validity of a particular instrument that purportedly measures a single attribute, a confirmatory one-factor model is often part of the methodological toolkit: if the model fits, this is taken as evidence for the construct validity of that particular instrument (e.g., physiological hyperarousal: Joiner et al., 1999; psychopathy: Patrick, Edens, Poythress, Lilienfeld, & Benning, 2006). Importantly, for quite a few scholars in applied psychology research, these latent attributes—which, in a latent variable model, are nothing more than mathematical abstractions in a set of equations—are (implicitly) endowed with a

realist meaning: for example, personality psychologists who believe that the Extraversion factor in the Big Five factor model exists in the minds of individual people and causes behavior (McCrae & Costa, 2008).

In earlier work, we have postulated a new theoretical framework in which psychological phenomena like MD are networks in which the symptoms are not caused by a latent attribute but, instead, are directly related (Borsboom, 2008b; Cramer, Borsboom, et al., 2012; Cramer et al., 2010). That is, for example, the symptoms insomnia and fatigue are not caused by the latent “MD” attribute but, instead, are directly linked: insomnia \(\rightarrow\) fatigue. Direct links need not be causal, they can be bidirectional as well (e.g., a bidirectional relation between depressed mood and concentration problems: depressed mood \(\leftrightarrow\) concentration problems). Such causal/bidirectional relations between symptoms make more sense: Why would one need a latent attribute to explain why not sleeping and being tired are highly correlated? A similar argument was made for intelligence, in which, for example, verbal and arithmetic skills are strongly correlated because these skills mutually influence one another, and not because both skills are caused by the same underlying intelligence attribute (van der Maas et al., 2006). Finally, why would one need a latent “extraversion” attribute to explain why people who like parties often like to be in the centre of attention (Cramer, van der Sluis, et al., 2012).

In a classic measurement model, the probability of a symptom to be present depends on the latent attribute: the more depressed one is, the higher the probability that depression symptoms will be endorsed. In a network of symptoms, the probability of symptom \(x\) to be present depends on whether the symptoms that are directly linked to symptom \(x\) were present at an earlier point in time: for example, the probability of having concentration problems at \(t = 1\) depends on whether sleep problems, fatigue, and worrying were present at \(t = 0\). Or, in the case of extraversion, the probability of liking parties depends on whether one likes people and likes being in the center of attention.

Now, suppose the network perspective paints an accurate picture of at least some psychological phenomena. What does this mean for current definitions of validity as defined in terms of attributes and measurement? There is no attribute, at least not an attribute as defined in the classic measurement model. This does, however, not invalidate current definitions of validity. Take the network in Figure E.1 as an example and suppose we want to know the probability of being tired (fati). As mentioned earlier, this probability depends on whether the other symptoms with which fatigue is linked are present in an earlier point in time: the more depressed mood (mood), insomnia (inso), and concentration problems (conc), the higher the probability of fatigue to become present later in time. As such, it makes sense to say that fatigue in this network is caused by depressed mood, insomnia, and concentration problems; likewise, it makes sense to say that insomnia, in this network, is caused by depressed mood, fatigue, and concentration problems and so on. That is, each symptom in this network is caused by the total activation of the other symptoms; and it makes perfect sense to define the total activation summed over all symptoms to be an attribute —not an attribute in the classical measurement sense (with realist connotation), but one that is a summary statistic for how symptoms are influenced by one another.

So, from a network perspective, an intelligence test does not measure latent \(g\) but it assesses the extent to which cognitive processes such as verbal abilities and arithmetic skills are activated/present. Likewise, a depression questionnaire does not measure a latent depression attribute, but it assesses the total amount of activation of symptoms in a depression network. But to what does this total activation amount? In other work, we have formulated the hypothesis that psychological disorders are complex systems that, governed by a set of (non)linear differential equations, move towards one or more attractors (Schmittmann et al., 2013). For example, the MD network moves towards one
of two attractors: a “depressed” and a “non-depressed” attractor. Whenever the MD network is, for example, in the “depressed” attractor, we say that the MD network is in a depressed state. And what we do when administering a depression questionnaire to someone is assess the state of the depression network: the more symptoms are present/active, the more the MD network is pushed towards a depressed state.

Does this novel formulation of attributes and measurements mean that validity can no longer be assessed with, for example, fitting a confirmatory factor model? No. One can fit a one-factor model to investigate a questionnaire that consists of, say, extraversion items. However, when the model fits, the interpretation should be different: no longer should one speak of “construct validity” in the sense that the fit is evidence of an extraversion construct (with a possible realist connotation) and measurable by the items in that particular questionnaire. Rather, from a network perspective, the interpretation should be that the fit provides evidence that these particular items have direct nontrivial links with one another: apparently, endorsing an item depends on endorsing other items. As such, the latent factor serves as the summary statistic of the summed total activation/presence of these items.

So what is validity then, when adhering to a network perspective? Departing from Newton’s as well as the Standards’ definition, validity should be about the degree to which an instrument measures what it purports to measure (Ruch, 1924; see also Borsboom, Cramer, et al., 2009). And, then, what do instruments generally purport to measure? From a network perspective, instruments are designed to assess whether a particular set of items form a network, that is, whether they are directly connected with one another. Although it might seem to be so, this perspective on validity does not radically depart from existing formulations. In fact, in many cases the outcome of a validation study will be the same - take for instance the items “23 + 1” and “feeling blue”. No matter the shape of the looking glass, “classic measurement”, or “network”, the conclusion will be that an instrument containing these two items is not a valid instrument for measuring MD.

However, the question of why this is so has different answers. From a classic measurement perspective, the instrument is not valid because there is no latent attribute that can explain the majority of covariation (if any, for the correlation between these items...
will be low) between the two items. That is, “23 + 1” does not measure MD. From a network perspective, the instrument is not valid because the two items are not directly linked: the probability of correctly answering “23 + 1” does not depend on endorsing “feeling blue” and vice versa. One will conclude that the item “23 + 1” is not part of the MD network. That is, the interpretational difference between the two perspectives lies in assuming a measurement (i.e., classic measurement: the items measure an attribute), versus a mereological relationship, between attributes and items: from a network perspective, the items are what constitutes an attribute. Thus, in sum, feeling blue and thinking about suicide are not measurements of depression; feeling blue and thinking about suicide is what being depressed is.