A Psychometric Network Perspective on the Validity and Validation of Personality Trait Questionnaires

ALEXANDER P. CHRISTENSEN[1]*, HUDSON GOLINO[2] and PAUL J. SILVIA[1]

[1]Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC USA
[2]Department of Psychology, University of Virginia, Charlottesville, VA USA

Abstract: This article reviews the causal implications of latent variable and psychometric network models for the validation of personality trait questionnaires. These models imply different data generating mechanisms that have important consequences for the validity and validation of questionnaires. From this review, we formalize a framework for assessing the evidence for the validity of questionnaires from the psychometric network perspective. We focus specifically on the structural phase of validation, where items are assessed for redundancy, dimensionality, and internal structure. In this discussion, we underline the importance of identifying unique personality components (i.e. an item or set of items that share a unique common cause) and representing the breadth of each trait's domain in personality networks. After, we argue that psychometric network models have measures that are statistically equivalent to factor models but we suggest that their substantive interpretations differ. Finally, we provide a novel measure of structural consistency, which provides complementary information to internal consistency measures. We close with future directions for how external validation can be executed using psychometric network models. © 2020 European Association of Personality Psychology

Key words: assessment; measurement; network analysis; personality scales and inventories; validity

INTRODUCTION

What are personality traits? Your answer likely implies certain hypotheses about the existence of traits and their underlying data generating mechanisms. These hypotheses are usually supported by your choice of psychometric model (Borsboom, 2006). Psychometric models come with a number of assumptions such as how traits cause variation in your measures and the meaning of scores derived from these measures (Borsboom, Cramer, Kievit, Scholten, and Franić, 2009; Cramer, 2012). These models also come with a number of other consequences such as considerations about how scales should be developed and validated.

For many researchers, personality traits are complex systems—that is, traits are *systems* in that they are composed of many components which interact with one another and *complex* in that their interactions with other systems are difficult to derive because of their dependencies and properties. Despite this view, personality traits are usually not modelled this way. Most psychometric models provide parsimonious perspectives on personality traits, which may arbitrarily carve joints into the fuzzy nature of personality. In addition, these models have causal implications that some researchers might not agree with. Therefore, there is a need for models that better align with how researchers think about personality.

One promising model comes from the emerging field of *network psychometrics* (Epskamp et al., 2018a). Psychometric network models have a simple representation: nodes (circles) represent variables (e.g. questionnaire items), and edges (lines) represent the unique associations (e.g. partial correlations) between nodes (Epskamp and Fried, 2018; Epskamp et al., 2018b). This representation supports the theoretical perspective, often referred to as the *network approach* (Borsboom, 2008, 2017), that psychological attributes are complex systems of observable behaviours that dynamically and mutually reinforce one another (Schmittmann et al., 2013).

From this perspective, personality traits resemble an emergent property of the interactions that occur between unique behavioural components—that is, traits are not any single component of the system but rather a feature of the system as a whole (Baumert et al., 2017; Cramer et al., 2012a). This suggests that traits emerge because some characteristics and processes within individual people tend to covary more than others (Mõttus and Allerhand, 2017), and when these relevant processes are aggregated, they reflect meaningful differences between people in the population (e.g. trait domains; Borkenau and Ostendorf, 1998; Cramer et al., 2012a).

Such an explanation of traits affords a novel context for how they should be assessed. First, it implies that behaviours that are associated with one trait may directly influence behaviours of another trait—the lines separating traits are more fuzzy than they are distinct (e.g. comorbidity in psychopathology; Cramer, Waldrop, van der Maas, and Borsboom, 2010). Second, there is an emphasis on a trait's components as much as the trait itself—a trait's observable components

*Correspondence to: Alexander P. Christensen, Department of Psychology, University of North Carolina at Greensboro, P.O. Box 26170, Greensboro, NC 27402-6170, USA.
Email: apchrist@uncg.edu

do not measure the trait but are instead part of the trait (Schmittmann et al., 2013). This suggests that a behaviour such as liking to go to parties is only one part of a causal collection of behaviours that we call extraversion (Borsboom, 2008; Cramer, 2012). This represents a refocusing on what *parts* of a trait are being measured rather than the premise that the trait *itself* is being measured. Finally, this explanation proposes that the behavioural components of a trait are *unique*, meaning they have distinct causes (Cramer et al., 2012a). This suggests that there should be a shift in how we model existing questionnaires where many related items are often used to measure a single attribute.

The intent of this paper is to elaborate on what the novel perspective provided by the network approach means for personality measurement and assessment. We focus specifically on the validity and validation of personality trait questionnaires with the goal of demonstrating how psychometric network models relate to modern psychometric perspectives. We place a particular emphasis on the structural analysis of validation (e.g. item analysis, dimension analysis, and internal structure; Flake, Pek, and Hehman, 2017; Loevinger, 1957). Before discussing validation, we first consider what it means for a questionnaire to be valid—a topic that has a defining role in the substantive interpretation of personality measures.

(TEST) VALIDITY OF PERSONALITY TRAIT QUESTIONNAIRES

The trait approach has a long tradition in personality, significantly shaping the last 30 years of research. Many contemporary theories of personality are inclined to accept traits as phenomena that exist in some form (e.g. concrete biological entities, abstract population summaries). Across theories, traits seek to provide parsimonious descriptions of broad between-person patterns of covariation at the population level (Baumert et al., 2017). Five or six higher order traits are commonly thought to represent the majority of these between-person differences, which are typically assessed using questionnaires (Lee and Ashton, 2004; McCrae and Costa, 1987). The validation of these questionnaires is a critical part of the research agenda (Flake et al., 2017).

There are many views on what validity means with the most common perspectives involving the interpretation of test scores (e.g. Cronbach and Meehl, 1955; Kane, 2013; Messick, 1995). This is not how we view validity; instead, we adopt the definition that validity refers to whether a test measures what it intends to measure (Borsboom et al., 2009; Cattell, 1946; Kelley, 1927). Borsboom and colleagues (2004) refer to this as *test validity*, which states that 'a test is valid for measuring an attribute if and only if (a) the attribute exists and (b) variations in the attribute causally produce variations in the outcomes of the measurement procedure' (p. 1061). An *attribute* refers to a property that exists prior to and independent of measurement (Loevinger, 1957). This definition of validity involves connecting the structure of an attribute to the response processes of a measure.

In this section, we provide an overview of what this definition means for the validity of personality trait questionnaires. Most of the heavy lifting for the relation between a personality trait and questionnaire is done by a researcher's choice of psychometric model (Borsboom, 2006). The most common model used for validation in the personality literature is the latent variable model (Flake et al., 2017). Therefore, we begin our discussion of validity by briefly reviewing how latent variable models make sense of the test validity criteria (for a more through treatment, see Borsboom, Mellenbergh, and van Heerden, 2003). We then move to psychometric network models, which have emerged as an alternative explanation for the coherence of traits. In terms of validity, much less has been put forward for psychometric network models and personality questionnaires; therefore, we spend most of this section reviewing its current state and discussing its meaning in the context of personality measurement. Throughout this section, we refer to a hypothetical questionnaire of extraversion to contextualize our points.

Latent variable perspective

In personality (and most of psychology), reflective latent variable models, where the indicators are regressed on the latent variable (i.e. causal arrows point from the latent variable to the indicators), are the standard conceptualization of measurement (Borsboom et al., 2003). A reflective latent variable model holds that the items in our questionnaire are a function of the latent variable, meaning that people's responses to our questionnaire are caused by their position on the latent variable (e.g. extraversion). Using this causal explanation, we can evaluate the criteria for test validity.

First, does the attribute extraversion exist? That is, does extraversion exist prior to and independently of our questionnaire? Many researchers would consider this question trivial; however, to maintain that extraversion is indeed an attribute that exists, then the latent variable must be causally responsible for the responses to our questionnaire (Borsboom et al., 2003). Indeed, this is how many researchers think about the relationship between personality traits and their questionnaires as well as what some theories of personality suggest (McCrae and Costa, 2008; McCrae et al., 2000).

The second criterion is the crux of test validity which, as Borsboom and colleagues (2004) point out, is not so straightforward. This is because it requires a theory for how extraversion can be linked to the response processes of our questionnaire. Difficulties arise because it's plausible (and even likely) that the processes that lead one person to respond with 'agree' to an item (e.g. 'I like to go to parties') and another person to respond 'agree' to the same item are different (Borsboom and Mellenbergh, 2007). An idealistic perspective is that people have different response processes but they select 'agree' on the same item because they are positioned similarly on extraversion. With this perspective, a common and implicit interpretation is that people possess some quantity of extraversion and it's the *difference* in these quantities that cause the variation in how people respond to our questionnaire. More simply, Alice scores higher on our

questionnaire than Bob because Alice is positioned higher on the extraversion continuum than Bob.

From a causal perspective, a defensible account for how extraversion causes variation in our questionnaire would be that 'population differences in position on [extraversion] cause population differences in the expectation of item responses' (Borsboom et al., 2003, p. 211). This implies that people in the population which occupy the same position on extraversion will typically respond similarly to the same items in our questionnaire. This brings us back to the first criterion: Does extraversion exist? Or rather, to what *extent* does extraversion exist? One claim would be that extraversion exists as a between-person attribute—that is, as a *population attribute*. A population attribute is not necessarily possessed by any one person in the population but rather represents between-person differences at the population level (Baumert et al., 2017; Cervone, 2005). This notion aligns well with the Allportian view that '[a] common trait is a category for classifying functionally equivalent forms of behaviour in a general population of people' (Allport, 1961, p. 347).

From a noncausal perspective, this means that extraversion could exist as a useful descriptor for comparing people rather than explaining their behaviours (Hogan and Foster, 2016; Pervin, 1994). Many personality researchers hold this view of reflective latent variables (e.g. Ashton and Lee, 2005; Goldberg, 1993). Therefore, researchers need not view a reflective latent variable as causal but rather as a summary statistic of the shared variance between items in our questionnaire. This leaves our questionnaire's validity as a subject of substantive interpretation—that is, it depends upon what researchers think they are measuring: traits as population level positions or traits as descriptive summaries of between-person differences in the population.

Psychometric network perspective

Psychometric network models have been proposed as an alternative explanation for the emergence of personality traits. From a psychometric network perspective, traits arise not because of a latent common cause but rather from the causal (bi)directional relationships between observed variables (Cramer et al., 2012a). This explanation suggests that latent traits are not necessary to explain how items in our questionnaire covary (Borsboom et al., 2009). Moreover, it implies that traits do not exist or at least they do not exist in a classical sense of measurement (i.e. causing variation in our questionnaire; Cramer, 2012). Instead, the relationship between extraversion and our questionnaire is a mereological one—that is, the items in our questionnaire do not measure extraversion but are part of it (Borsboom, 2008; Cramer et al., 2012a).

Extraversion is therefore a summary statistic for how personality components are influenced by one another (e.g. liking to talk to people → liking to go to parties ↔ liking to meet new people; Cramer, 2012). In this sense, extraversion exists as a state of the network or the stable organization of dynamic personality components that are mutually activating one another (Cramer et al., 2012a; Schmittmann et al., 2013).

Our questionnaire thus refers to the state of a specific set of personality components that are causally dependent on one another and form a network (Cramer, 2012). The state of the network is determined by the total activation of these components and is what we refer to as extraversion—that is, the more personality components that are active, the more the network is pushed towards an extraverted state (Cramer, 2012).

In the context of a between-person network model, our questionnaire's network represents the aggregation of the average activation of each component across *within-person networks* (i.e. each individual person's network across several time points; Cramer et al., 2012a; Epskamp et al., 2018b). Both theory and empirical evidence appear to support this claim. From a Whole Trait Theory perspective (Fleeson and Jayawickreme, 2015), people's responses to items in self-report questionnaires correspond to their locations and maximums of their density distributions for the respective within-person states (Fleeson, 2001; Fleeson and Gallagher, 2009). When aggregated, these states tend to correspond to self-reported traits (Rauthmann, Horstmann, and Sherman, 2019), and when compared across people, these states typically produce the between-person traits (Borkenau and Ostendorf, 1998; Hamaker, Nesselroade, and Molenaar, 2007). This interpretation leaves open an important question: What then do personality components refer to?

Personality components
Cramer and colleagues (2012a) define personality components as 'every feeling, thought, or act' that is associated with a 'unique causal system' (p. 415). In most instances, these components refer to items of a questionnaire. A key point of emphasis in their definition is that these components are *unique* in that they are causally autonomous (i.e. distinct causal processes). For many existing personality questionnaires, items are often not unique. Instead, facets or narrower characteristics of a trait (e.g. gregariousness, warmth, and assertiveness) are composed of many closely related and sometimes redundant items. In this sense, personality components that comprise extraversion may be items but they also may be facets (Costantini and Perugini, 2012).

In our view, this represents a key difference between facets and components: *Facets* are a collection of related items (not necessarily sharing a unique common cause), while *components* are an item or set of items that share a unique common cause. This distinction is important because some facets in existing questionnaires reflect a homogeneous cause (and are therefore considered a component), while other facets reflect heterogeneous causes, which must be separated into unique components.[1] Therefore, a facet from the network perspective would be a set of unique components that coalesce into a meaningful suborganization of a trait's domain. From this perspective, it becomes imperative that researchers determine whether items and facets of an existing questionnaire are distinct autonomous causal components

---

[1]The use of 'reflect' in our language is on purpose and implies a common cause that can be associated with a reflective latent variable.

or if they reflect a common cause (Hallquist, Wright, and Molenaar, 2019).

Based on this definition, personality components appear to closely resemble attributes. In this way, extraversion may exist as a *composite attribute* or an attribute that is composed of many other attributes (Borsboom and Mellenbergh, 2007). The number of attributes that constitute extraversion then becomes a function of the sampling properties from its domain of representative attributes (McDonald, 2003). Importantly, the selection of attributes will change the composition of the network, meaning that different questionnaires will have different compositions despite still plausibly referring to extraversion (Markus and Borsboom, 2013). Extraversion should then be viewed as a *finite* universe of attributes where there are a limited number of unique attributes that can comprise it (McDonald, 2003). Therefore, there is a particular need to identify and validate the content of each personality trait's domain (Markus and Borsboom, 2013).

It's tempting to say that researchers should only measure attributes that represent one domain; however, it's unlikely that attributes of a personality trait will exist independently of other trait domains (Schmittmann et al., 2013; Schwaba, Rhemtulla, Hopwood, and Bleidorn, 2020; Sočan, 2000). An item like 'enjoys talking to people,' for example, certainly represents the domain of extraversion, but it may also represent the domain of agreeableness. Common examples of this cross-domain entanglement often occur in psychopathological comorbidity (Cramer et al., 2010). Thus, distinctions between what attributes constitute the extraversion domain become rather fuzzy and a matter of degree because of the overlap attributes can have with other domains (Schmittmann et al., 2013). Such fuzziness is likely common in personality where attributes may not clearly delineate between where one trait begins and another ends (Connelly, Ones, Davies, and Birkland, 2014). Indeed, this is exactly what functionalist and complex system theories of personality suggest (Cramer et al., 2012a; Perugini, Costantini, Hughes, and De Houwer, 2016; Read et al., 2010; Wood, Gardner, and Harms, 2015) and what recent psychometric network analyses of personality traits find (Schwaba et al., 2020).

Validity from the network perspective
This leads to the question of how extraversion (as a composite attribute) can cause variation in our questionnaire. Quite simply, it does not: There is no link between the (composite) attribute and the questionnaire's response processes because it does not exist (Schmittmann et al., 2013). This is because no single attribute that extraversion is composed of will directly assess extraversion; instead, each attribute assesses parts of the extraversion domain (Borsboom, 2008; Borsboom and Mellenbergh, 2007; Cramer et al., 2012b). With this perspective, we can say that the variation in our questionnaire arises from the sampling of attributes in the representative domain (Borsboom and Mellenbergh, 2007), which is clear from studies that have examined several different questionnaires (e.g. Christensen et al., 2019; Schwaba et al., 2020).

When evaluating whether our questionnaire is a valid measure of extraversion from the network perspective, we must shift the evaluation from the validity of extraversion as an attribute to the validity of its components. This does not rule out the validity of the questionnaire but rather shifts the perspective such that our questionnaire is measuring the state of the network composed of causal connected components that we refer to as extraversion. The explanation for the variation of our measurement thus comes from the reciprocal cause and effect of other attributes in the network.

This explanation does not come without consequence. The issue of connecting the attribute to response processes is merely side stepped from personality traits to personality components. The response processes in network models are assumed to lie in the reciprocal cause and effect of other components. Indirectly, this suggests that the responses processes of one component has reciprocal causes and effects on processes of other components. This point, however, is circular in that it still does not specify what the response processes are.

Although the network perspective avoids introducing latent variables to account for these response processes, it does not avoid the question of how they occur. To this end, it is important for personality researchers from the network perspective to connect response processes to personality components. More specifically, researchers must seek to specify how response processes of one component can cause and effect processes of another component. We do not claim to have a definitive answer to this issue but highlight it as one that is particularly perplexing and requires sophisticated research designs (e.g. Costantini, Richetin, et al., 2015).

VALIDATION OF TRAIT QUESTIONNAIRES FROM A PSYCHOMETRIC NETWORK PERSPECTIVE

Our discussion of validity to this point has been about whether a questionnaire possesses the property of being valid. This discussion sets up how psychometric evaluations of a questionnaire should be substantively interpreted during validation. Validation differs from validity in that it is an ongoing activity which seeks to describe, classify, and evaluate the degree that empirical evidence and theoretical rationales support the validity of the questionnaire (Borsboom et al., 2004; Messick, 1989). Validation usually entails three phases: substantive, structural, and external (Flake et al., 2017). Our main focus will be on the structural phase, which primarily consists of establishing evidence that our questionnaire measures what we intend it to through item, dimension, and internal structure (e.g. internal consistency) analyses.

Validation from a psychometric network perspective has received relatively little attention. To date, psychometric network models have mainly been used as a novel measurement tool, which has led to an alternative account for the formation of traits (Costantini, Epskamp, et al., 2015; Cramer et al., 2012a). When it comes to psychometric assessment, the scope of psychometric networks has been far more limited (e.g. dimension reduction methods; Golino and Epskamp, 2017; Golino et al., 2020). There does, however, appear to

be some potential because networks have been shown to be mathematically equivalent to latent variable models (Guttman, 1953; Kruis and Maris, 2016; Marsman et al., 2018).

The key to distinguishing network psychometrics models from latent variable models is to establish how the measures of these models differ in their substantive interpretations (i.e. hypothesized data generating mechanisms; van Bork et al., 2019). We will draw on several points from the previous section on validity to elaborate on these interpretations. In the end, the aim of this section is to take the initial steps towards a formalized framework for the use of psychometric network models in the validation of personality questionnaires.

Overview

To achieve this aim, we divide this section into three parts, which represent the order in which researchers should proceed with structural validation from the psychometric network perspective. First, we cover the initial phase of redundancy analysis for reducing redundancy in personality questionnaires. Next, we discuss dimension analyses. Within this section, we connect communities and node strength of network models to factors and factor loadings of latent variable models, respectively. Finally, we present a novel measure of internal structure that can be used to assess the extent to which a scale (or dimension) is composed of a set items that are homogeneous *and* interrelated in a multidimensional context.[2]

In our discussion, it's important that we make clear that we view network models as complementary to latent variable models and therefore suggest that they can be synergistically leveraged. The main difference between them, as we discussed in our section on validity, is the proposed data generating mechanisms. From a statistical point of view, network models offer additional information about the relationship between variables that are not possible in latent variable models since in the latter the relationship between items are accounted by the factor. Analyses about the structure of the system (e.g. topological analysis), for example, can be implemented in network models, helping researchers uncover important aspects of the system (Borsboom, Cramer, Schmittmann, Epskamp, and Waldorp, 2011). Because of this, we connect network models to latent variable models (where applicable) and highlight the substantive differences that these models imply. As a general point, we recommend at least 500 cases when performing these analyses, which is based on previous simulation studies (Christensen, 2020; Golino et al., 2020).

Redundancy analysis

In scale development, a researcher must establish what items to include, which involves determining the desired specificity

and breadth of the trait(s) the researcher is trying to measure. Greater specificity leads to scales that have higher internal consistency, which increases the likelihood that the researcher is measuring the same attribute while reducing idiosyncrasies specific to each item (DeVellis, 2017). Greater breadth leads to scales that have higher item-specific variance, which increases the coverage of the representative domain (McCrae, 2015). In many existing trait questionnaires, researchers have focused on achieving a balance of both—that is, some facets reflect a single narrow attribute, while other facets are composites of several attributes.

One recent suggestion for questionnaires aimed at trait domains is to favour breadth in order to maximize information and efficiency (McCrae and Mõttus, 2019). Based on what we've outlined in our section on validity, psychometric network models align well with this suggestion. Indeed, a key notion of network psychometrics is that personality traits are composed of unique causal components, meaning the components are not exchangeable with other components of the system (Cramer et al., 2012a). As a consequence, these components should be unique rather than redundant to reduce latent confounding (Hallquist et al., 2019). This implication perhaps marks the biggest validation difference between network and latent variable models.

Because most existing personality scales have been developed from a latent variable perspective, researchers must make careful considerations about using psychometric network models with existing scales because they are likely to have homogeneous facets (Costantini and Perugini, 2012). Take, for example, the SAPA Personality Inventory (Condon, 2018) where items 'Hate being the center of attention', 'Make myself the center of attention', 'Like to attract attention', and 'Dislike being the center of attention' clearly have a common underlying attribute: attention seeking. From the psychometric networks perspective, these items are not unique components themselves but comprise a single unique component. This makes the first step of questionnaire validation from a psychometric network perspective to identify and handle redundancy in scales.

An approach to statistically identify redundancy

In the literature, the network measure, *clustering coefficient*, has been considered as a measure of redundancy in personality networks (Costantini et al., 2019; Dinić, Wertag, Tomašević, and Sokolovska, 2020). A node's clustering coefficient is the extent to which its neighbours are connected to each other, forming a triangle. Although this measure is useful for describing whether a node is locally redundant, it does not provide information about which nodes in particular a target node is redundant with. Here, we conceptually describe an approach to identify whether a node is statistically redundant with other nodes in a network.

Our approach begins by first computing a similarity measure between nodes. One method for doing so is called *weighted topological overlap* (Zhang and Horvath, 2005), which quantifies how similar two nodes' connections to other nodes are. More specifically, it quantifies the similarity between the magnitude and direction of two nodes' connections to all other nodes in the network. In biological networks,

---

these measures have been used to identify genes or proteins that may have a similar biological pathway or function (Nowick, Gernat, Almaas, and Stubbs, 2009). Thus, greater topological overlap suggests that two genes may belong to the same functional class compared to those with less overlap. In the context of a personality network, nodes that have large topological overlap are likely to have shared functional or latent influence. From a more traditional psychometrics perspective, one method would be to identify items that have high residual correlations after the variance of facets and factors have been removed.[3]

Although the weighted topological overlap measure provides numerical values, from no overlap (0) to perfect overlap (1), for each node pair in the network, it does not include a test for significance. In order to determine which node pairs overlap significantly with one another, we apply the following approach. First, we obtain only the values that are greater than zero—node pairs that have a topological overlap of zero are not connected in the network and are therefore not informative for determining significance of overlap. Next, we fit a distribution to these non-zero values using the *fitdistrplus* package (Delignette-Muller and Dutang, 2015) in R. The parameters (e.g. μ and σ from a normal distribution) from the best fitting distribution (based on Akaike information criterion) are then used as our probability distribution.

For each node pair with a non-zero topological overlap value, we compute the probability of achieving its corresponding value from this distribution. These probabilities correspond to p values. Using a multiple comparison method, node pairs whose p values are less than the corrected alpha are considered to be significantly redundant. We've implemented this approach in the *EGAnet* package (Golino and Christensen, 2020) in R under the function `node.redundant` (see the Node Redundancy section in the Supporting Information). Results from one simulation study found that the adaptive alpha multiple comparison correction method (Pérez and Pericchi, 2014) had the fewest false positives, false negatives, and highest accuracy of all the methods tested (Christensen, 2020).

Options for handling redundant nodes
This approach provides quantitative evidence for whether certain items are redundant. We recommend, however, that researchers verify these redundancies and use theory to determine whether two or more items represent a single attribute (i.e. a common cause). There are two options that researchers can take when deciding how to handle redundancy in their questionnaire. The more involved option is to remove all but one item from the questionnaire. When taking this option, there are a few considerations researchers must make. Qualitatively, which item represents the most general case of the attribute? Often items are written with certain situations attached to them (e.g. 'I often express my opinions in group meetings'; Lee and Ashton, 2018), which may not apply to all people taking the questionnaire. Therefore, more general items may be better because they do not represent a

situation-specific component of an attribute (e.g. 'I often express my opinions'). Quantitatively, which item has the most variance? This is a common criterion in traditional psychometrics because greater variation suggests that this item better discriminates between people on the specific attribute (DeVellis, 2017).

The more straightforward option is to combine items into a single variable. This can be done by estimating a reflective latent variable consisting of the redundant items and using the latent scores. We strongly recommend this latter approach because it retains all possible information from available data. We've implemented an interface to manage this second option using the `node.redundant.combine` function in the *EGAnet* package in R. We describe how to apply this approach in the supporting information, including heuristics to use when deciding which items are redundant.

Dimension analysis

Dimensionality assessment is an integral step for validating the structure of a questionnaire. The general consensus among researchers is that personality traits are hierarchically organized at different levels of breadth and depth (John and Srivastava, 1999; McCrae and Costa, 2008). Usually, trait domains are decomposed into facets, which are further broken down into items (McCrae and Costa, 1987, 2008). More recently, aspects (between traits and facets) were added to the hierarchy (DeYoung, Quilty, and Peterson, 2007). Personality questionnaires tend to follow this structure with most assessing multiple domains and facets—there typically are five or six trait domains and for every domain, there are several facets (ranging from two to nine).

In traditional psychometrics, factor models [e.g. exploratory factor analysis (EFA)] are the most common method used to assess the dimensionality of a trait domain (Flake et al., 2017). In psychometric networks, the main methods used to assess dimensionality of the network are called community detection algorithms (Fortunato, 2010). These algorithms identify the number of communities (i.e. dimensions) in the network by maximizing the connections within a set of nodes while minimizing the connections from the same set of nodes to other sets of nodes in the network. Rather than these communities forming because of a common cause, psychometric network models suggest that dimensions emerge from densely connected sets of nodes that form coherent subnetworks within the overall network.

Despite these frameworks proposing different data generating mechanisms, the data structures do not necessarily differ (van Bork et al., 2019). Indeed, a researcher can fit a factor model to a data structure generated from a network model with good model fit (van der Maas et al., 2006). Similarly, a network model with a community detection algorithm can be fit to a data structure generated from a factor model and identify factors (Golino and Epskamp, 2017; Golino et al., 2020). This underlying equivalence follows from the fact that any covariance matrix can be represented as a latent variable and network model (van Bork et al., 2019). The statistical equivalence between these models has been well documented (e.g. Epskamp et al., 2018a; Guttman,

---

[3]We thank the anonymous reviewer who pointed out this possibility.

1953; Kruis and Maris, 2016; Marsman et al., 2018). Therefore, factors of a latent variable model and communities of a network model are statistically equivalent (Golino and Epskamp, 2017).

Indeed, Guttman (1953) demonstrates that there is a direct equivalence between network and factor models. Although network models were not yet specified in the area of psychology, Guttman (1953) proposed a new factor analytic approach termed *image structural analysis*, which is essentially a network model with node-wise estimation using multiple regression (e.g. Haslbeck and Waldorp, 2015). Guttman mathematically demonstrated how image structural analysis relates to factor models and suggested that factor models were a special case of the node-wise network model where the errors of the variables are made to be orthogonal.[4] Therefore, the difference between the models is their suggested data generating mechanisms, which is provided by their visual representations.

It's important that we acknowledge that in some cases, factors of a factor model may represent causally dependent interactions between components (rather than a common cause); in other cases, communities of a network model may represent a common cause (rather than causally dependent interactions between components). Moreover, other explanations could be that external causes such as situational factors (Cramer et al., 2012a; Rauthmann and Sherman, 2018) or goals and motivations (Read et al., 2010) could lead to personality dimensions.

Exploratory graph analysis
The most extensive work on dimensionality in the psychometric network literature has been with a technique called *exploratory graph analysis* (EGA; Golino and Epskamp, 2017; Golino et al., 2020). The EGA algorithm works by first estimating a Gaussian graphical model (Lauritzen, 1996), using the graphical least absolute shrinkage and selection operator (GLASSO; Friedman, Hastie, and Tibshirani, 2008), where edges represent (regularized) partial correlations between nodes after conditioning on all other nodes in the network. Then, EGA applies the Walktrap community detection algorithm (Pons and Latapy, 2006), which uses random walks to determine the number and content of communities in the network (see Golino et al., 2020, for a more detailed explanation).[5] Several simulation studies have shown that EGA has comparable or better accuracy for identifying the number of dimensions than the most accurate factor analytic techniques (e.g. parallel analysis; Christensen, 2020; Golino and Demetriou, 2017; Golino and Epskamp, 2017; Golino et al., 2020).

Beyond performance, EGA has several advantages over more traditional methods. First, EGA does not require a rotation method. Rotations are rarely discussed in the validation

(e.g. estimation of factor loadings; Browne, 2001; Sass and Schmitt, 2010). For EGA, orthogonal dimensions are depicted with few or no connections between items of one dimension and items of another dimension. Second, researchers do not need to decide on item allocation—the algorithm places items into dimensions without the researcher's direction. For EFA, in contrast, researchers must decipher a factor loading matrix. Third, the network plot depicts some dimensions as more central than others in the network (see Dimensionality section in the Supporting Information). Thus, EGA can be used as a tool for researchers to evaluate whether the items of their questionnaire are coalescing into the dimensions they intended and whether the organization of the trait's structure is what they intended. Finally, the network plot also depicts levels of a trait's hierarchy as continuous—that is, items can connect between different facets and traits. This supports a fuzzy interpretation of the trait hierarchy where the boundaries between items, facets, and traits are blurred.

With these advantages, it's important to note their similarities to factor analytic methods. For instance, most community detection algorithms used in the literature (including the Walktrap) sort items into single dimensions. This creates a structure that is akin to a typical confirmatory factor analysis (CFA; i.e. items belonging to a single dimension), which constrains the interpretation of a continuous hierarchy. There are, however, algorithms in the broader network literature that allow for overlapping community membership (e.g. Blanken et al., 2018), which may better represent these fuzzy boundaries and how researchers think about personality. Another limitation is that the factor loading matrix of an EFA model can equivalently represent the complexity of items relating to other items and loading onto other dimensions. Network models, however, provide intuitive depictions of these interactions (Bringmann and Eronen, 2018). Therefore, even though EFA loading matrices represent this complexity, it requires a certain level of psychometric expertise for a researcher to intuitively view the matrix this way. Moreover, network plots can reveal exactly which items are responsible for the cross-domain relationships in a way that an EFA loading matrix cannot.

Loadings
Recent simulation efforts, however, have demonstrated that network models can be used to estimate an EFA loading matrix equivalent. In a series of simulation studies, Hallquist, Wright, and Molenaar (2019) demonstrated that the network measure, *node strength* (i.e. the sum of a node's connections), was roughly redundant with CFA factor loadings. A notable finding in one of their studies was that a node's strength could potentially be a blend of connections within and between dimensions. Based on this result, they suggested that researchers should reduce the latent confounding of the network measure to avoid misrepresenting the relationships between components in the network.

Heeding this call, Christensen and Golino (2020) derived a measure called *network loadings*, which represents the standardization of node strength split between dimensions. More specifically, a node's strength was computed for only

---

[4] We thank Denny Borsboom for pointing us to the Guttman's (1953) paper.
[5] A recent simulation study used the EGA approach and examined different community detection algorithms, finding that the Louvain (Blondel, Guillaume, Lambiotte, and Lefebvre, 2008) and Walktrap algorithms were the most accurate and least biased of the eight algorithms tested (including two parallel analysis methods; Christensen, 2020).

the connections it had to other nodes in each dimension of the network. They demonstrated that these network loadings could effectively estimate the simulated population (or true) loadings. Moreover, they found that network loadings more closely resembled EFA loadings but also had some loadings of zero like CFA loadings. This suggests that the network loadings represent a complex structure that is between a saturated (EFA) and simple structure (CFA). In sum, they suggest that these network loadings can be used as an equivalent to factor loadings (e.g. see Table SI3).

Although these metrics are statistically redundant, they arguably differ in a substantive way. Factor loadings suggest that items 'load' onto factors, which is provided by items being regressed on the factors. If interpreted in a substantive way, they represent how well one observable indicator is related to the factor—that is, how well an item represents or measures the latent factor. The substantive interpretation of node strength does not suggest this, however, they may be epistemologically related. From a substantive standpoint, we argue that these network loadings represent each node's contribution to the emergence of a coherent dimension in the network. In this sense, we can connect the substantive meanings of network and factor loadings: the more one item contributes to a dimension's coherence, the more the item reflects the underlying dimension. A researcher's substantive interpretation will favor one interpretation over the other, but ultimately, they statistically resolve to roughly the same thing (Christensen and Golino, 2020; Guttman, 1953; Hallquist et al., 2019).

Internal structure analysis

Internal consistency
Analyses that quantify the internal structure of questionnaires have been dominated by internal consistency measures, which are almost always measured with Cronbach's α (Cronbach, 1951; Flake et al., 2017; Hubley, Zhu, Sasaki, and Gadermann, 2014). In a review of 50 validation studies randomly selected from *Psychological Assessment* and *European Journal of Personality Assessment* during the years 2011 and 2012, α was reported in 90% and 100% of the articles, respectively (Hubley et al., 2014). Similar numbers were obtained in a review of 35 studies in the *Journal of Personality and Social Psychology* during 2014, with 79% of scales that included two or more items ($n = 301$) reporting α (Flake et al., 2017). More often than not, α was the sole measure of structural validation. In short, the use of α in validation is pervasive (McNeish, 2018).

Despite α's prevalence, there are some serious issues (Dunn, Baguley, and Brunsden, 2014; Sijtsma, 2009). These issues range from improper assumptions about the data (e.g. τ equivalent vs. congeneric models; Dunn et al., 2014; McNeish, 2018) to misconceptions about what it actually measures (Schmitt, 1996; Sijtsma, 2009). Although newer internal consistency measures (e.g. ω; Dunn et al., 2014; McDonald, 1999; Zinbarg, Yovel, Revelle, and McDonald, 2006) account for improper assumptions about the data, misconceptions about internal consistency still abound. One of the more persistent misconceptions is that internal

consistency measures assess unidimensionality (Flake et al., 2017). This misconception likely stems from confusion over the difference between *internal consistency* (the extent to which items are interrelated) and *homogeneity* (a set of items that have a common cause; Green, Lissitz, and Mulaik, 1977). Based on these definitions, internal consistency is necessary but not sufficient for homogeneity (Schmitt, 1996).

We believe that many misconceptions arise because there is a mismatch between what researchers intend to measure and what they are actually measuring. Much like validity, the psychometric concept of internal consistency seems divorced from how researchers think about it (Borsboom et al., 2004). This is because most researchers know that the items of their scale are interrelated—they were designed that way. When framed in this light, internal consistency measures are more of a 'sanity check' than an informative measure. To better understand what researchers intend to measure, we can look at how they use these measures: Researchers use them to validate the consistency of the structure of their scales (Flake et al., 2017). That is, researchers use them to know whether their scale's structure is consistent which *implies* internal consistency and *assumes* homogeneity.

From a latent variable perspective, the solution is straightforward: test if a unidimensional model fits and compute an internal consistency measure (Flake et al., 2017; Green et al., 1977). From a psychometric network perspective, this is not the case. First, there is an inherent incompatibility with computing an internal consistency measure from the network perspective. Internal consistency measures are typically a variant on the ratio between the common covariance between items and the variance of those items (McNeish, 2018). In the estimation of networks, most of the common covariance is removed, leaving only the correlations between item-specific variance (Forbes, Wright, Markon, and Krueger, 2017, 2019).

Second, scales and their items in networks are interrelated, usually with cross-connections occurring throughout. This is more than likely to be true for personality scales (Sočan, 2000). Therefore, it's important to know whether a set of items are causally dependent and form a unidimensional network but also whether they remain as a coherent subnetwork nested in the rest of the network. Said differently, questionnaires often contain scales that are assumed to be unidimensional but it's unclear whether these scales remain unidimensional when other items and scales are added (i.e. in a 'multidimensional context'). Therefore, internal consistency measures do not capture whether scales (or dimensions) remain unidimensional within the context of other items and dimensions. Regardless of psychometric model, this seems to be a more informative measure of what most researchers want to know and say about their scales—that they are unidimensional and internally consistent.

Structural consistency
We refer to this measure as *structural consistency*, which we substantively define as the extent to which causally coupled components form a coherent subnetwork within a network. Using extant terminology, structural consistency is the extent

to which items in a dimension are homogeneous *and* interrelated given the multidimensional structure of the questionnaire. In other words, it is the combination of homogeneity and internal consistency in a multidimensional context. We view the inclusion of other dimensions as a particularly important conceptual feature because a dimension could have high homogeneity and internal consistency but when placed in the context of other related dimensions it's structure falls apart (i.e. it is no longer unidimensional). This renders the interpretation of that dimension in a multidimensional context relatively ambiguous even when its interpretation is clear in a unidimensional context (i.e. examined in isolation).

A recently developed approach called *bootstrap exploratory graph analysis* (bootEGA; Christensen and Golino, 2019) can be used to estimate this measure. bootEGA applies a *parametric* and *non-parametric* bootstrap approach, but for the structural consistency measure, we focus on the parametric approach. The parametric approach begins by estimating a GLASSO network from the data and taking the inverse of the network to derive a covariance matrix. This covariance matrix is then used to simulate data with the same number of cases as the original data from a multivariate normal distribution.

EGA is then applied to this replicate data, obtaining each item's assigned dimension. This procedure is repeated until the desired number of samples is achieved (e.g. 500 samples).[6] The result from this procedure is a sampling distribution for the total number of dimensions and each item's dimension allocation. Although a number of statistics can be computed, we focus on two: structural consistency and item stability. To derive both statistics, the original EGA results (i.e. empirically derived dimensions) are used.

Structural consistency is derived by computing the proportion of times that each empirically derived dimension is *exactly* (i.e. identical item composition) recovered from the replicate bootstrap samples. If a scale is unidimensional, then structural consistency reduces to the extent that the items in the scale form a single dimension—that is, the proportion of replicate samples that also return one dimension. The range that structural consistency can take is from 0 to 1. A dimension's structural consistency can only be 1 if the items in the empirically derived dimension conform to that dimension across all replicate samples. Such a measure leads to an important question: What's happening when a dimension is structurally *inconsistent*?

To answer this, *item stability* or the proportion of times that each item is identified in each empirically derived dimension across the replicate samples can be computed. This relatively simple measure not only provides insight into which items may be causing structural inconsistency but also the other dimension(s) these items are being placed in. On the one hand, two items of our hypothetical dimension might be at the root of the structural inconsistency; on the other hand, it might be multiple items are at the root of the structural inconsistency. In either case, examining each item's

replication proportions across dimensions can reveal whether they are forming a new separate dimension (only replicating in a new dimension), fit better with another dimension (replicating more with another dimension), or identify as multidimensional items (replicating equally across multiple dimensions). The latter two explanations can be verified using the network loading matrix. An example of these analyses are provided in the Structural Consistency section of the Supplementary Information.

In practice, the goal of structural consistency is to determine the extent in which a dimension is composed of a set items that are homogeneous and interrelated in the context of other dimensions. The importance that a dimension of a questionnaire has a high structural consistency is up to the researcher's intent. For many dimensions in personality, items may be multidimensional, which will lead to lower values of structural consistency. Therefore, lower structural consistency is not a bad thing if it is what the researcher intends. More importantly, the items that are leading to the lower structural consistency can be identified with item stability statistics, which may help researchers decided whether an item is multidimensional or fits better with another dimension. At this point, it is too early to make recommendations for what 'high' or 'acceptable' structural consistency means. Ultimately, simulation studies are necessary to develop such standards.

DISCUSSION

Questionnaires have been and will likely continue to be a standard format for the measurement of personality attributes. The validity of what questionnaires claim to measure, however, has rarely been explicated in the contemporary personality literature. Instead, psychometric models have been applied without much consideration of their causal implications. In this paper, we provided a review on the validity of personality trait questionnaires from the latent variable and psychometric network perspectives. The goal of our review was not to argue for one approach more than another but to elaborate on how questionnaires can be viewed as valid measures of personality traits or attributes. These views imply different substantive interpretations about the underlying data generating mechanisms, which are important for understanding the meaning of what's being measured and how psychometric measures substantively inform our measurement.

In our review, we took special interest in elaborating on the psychometric network perspective because few articles have focused on their measurement when applied to personality questionnaires. Much like latent variable models, psychometric network models have been readily applied by researchers without much consideration about their causal implications or how network measures should be interpreted in a psychological context (Bringmann et al., 2019). Based on our review, we propose a substantive interpretation of node strength (network loadings) that is appropriate in the context of dimensions and overall network—that is, a node's contribution to the emergence of a coherent subnetwork or

---

[6]A total of 500 replicate samples should be an adequate number to achieve an accurate estimate; however, researchers can increase this number to obtain greater precision.

network. This interpretation is by no means definite; however, we believe that it is a more appropriate interpretation than what has been put forward in the literature.

More specific to personality, we explicated an initial framework for how psychometric networks can be used to validate the structure of personality trait questionnaires. One point of emphasis was on reducing the redundancy of components in the network. This is because components of a network are defined as 'unique' and 'causally autonomous' (Cramer et al., 2012a). We described a novel approach to detect an item's redundancy with other items in the network, which can aid researchers in this endeavour. Moreover, we provided some general recommendations for removing or combining redundant items. Following from this emphasis, greater exploration of the unique components that represent the domain of personality traits is necessary so that a specific set of attributes can be defined. This is unlikely to be an easy task because personality traits are multifaceted and interrelated, which suggests that representation of a domain may be a matter of degree rather than clear cut definitions (Schmittmann et al., 2013; Schwaba et al., 2020).

This puts determining appropriate coverage of each trait's domain at the forefront of the psychometric network research agenda in personality. Indeed, determining appropriate coverage of each trait's domain is still an active area of research and requires more attention than it's been given in the past. In many cases, this will require sampling from attributes that may lie just outside of a trait's domain. We recommend that researchers focus more on the extent to which attributes represent each domain rather than assuming an existing questionnaire's domain coverage is sufficient. One place to start is by examining the unique items of several personality questionnaires in a single network domain (e.g. Christensen et al., 2019). Multiple domains and outcome measures could also be included to help determine these boundaries (e.g. Afzali, Stewart, Séguin, and Conrod, 2020; Costantini, Richetin, et al., 2015; Schwaba et al., 2020).

When it comes to item and dimension analyses, many of the statistics for latent variable models (i.e. factor loadings and dimensions) are mathematically equivalent to psychometric network models (Christensen, 2020; Golino and Epskamp, 2017; Hallquist et al., 2019). We argued that the key difference between these models is their substantive interpretations, which suggest very different data generating mechanisms (van Bork et al., 2019). At this point, disentangling these models is a nascent area of research.

Kan, van der Maas, and Levine (2019), for example, show how fit indices can be applied to network models so that they can be compared to confirmatory factor analysis models. They also demonstrated how the comparison of networks over groups could be achieved (see also Epskamp, 2019). van Bork et al. (2019) developed an approach to compare the likelihood that data were generated from a unidimensional factor model or sparse network model by assessing the proportion of partial correlations that have a different sign than the corresponding zero-order correlations and the proportion of partial correlations that are stronger than the corresponding zero-order correlations (greater proportions for both increase the likelihood of the sparse network model). Approaches like these can be used to determine whether a latent variable or psychometric network model may be more appropriate for the data. Ultimately, we believe that the choice of model will not significantly affect the outcomes of these dimension-related analyses.

Finally, we introduced a novel measure, structural consistency, to quantify a questionnaire's internal structure. Part of the motivation for this measure was the need to move beyond measures of internal consistency, which we believe do not necessarily align with what researchers intend to measure. Notably, we do not view this measure as incompatible with internal consistency but rather complementary. As we discussed, a dimension could be homogeneous (i.e. unidimensional) and internally consistent (i.e. interrelated) but it may not remain homogeneous in a multidimensional context. Such a condition is likely to occur in personality measures where components of traits tend to be interrelated. In general, this measure adds to the internal structure methods that researchers can use for validating the structure of their questionnaire.

Steps towards external validation

To this point, we've described our conceptual framework for the structural validation of personality questionnaires from the network perspective. This framework leaves open questions related to external validation. How, for example, do outcome variables relate to the components in personality networks? What about covariates? How does this fit with contemporary trends for evaluating the unique predictive value of items? Moreover, what if the researcher is interested in relating the trait itself to outcomes rather than components? We briefly discuss these questions in turn.

Personality–outcome relations are a fundamental part of personality research and the validation of personality assessment instruments. These relations are just as fundamental to the network perspective as more traditional perspectives. Our suggestion for this is relatively simple: include the outcome(s) of interest in the network. Similarly, important covariates should also be added to the network. Indeed, Costantini, Richetin et al. (2015) used this approach to evaluate how facets of conscientiousness were related to measures of self-control, working memory, self-report behaviours related to conscientiousness, and implicit attitudes of conscientiousness descriptors. Similarly, Afzali et al. (2020) longitudinally examined items of the Substance Use Risk Profile Scale and their relations to cannabis and alcohol use in adolescences. These studies not only provide a more complex evaluation of the relations between personality and outcomes but also provide more targeted item generation for future measures (e.g. including more items measuring transgression in sensation-seeking personality indicators; Afzali et al., 2020).

We propose that because networks are often estimated using the GLASSO approach researchers can interpret the partial correlation coefficients between outcomes and components in the network as if they were entered into a regularized regression. Regularized regression has already

been effectively used in the literature to evaluate personality–outcome relations (Mõttus, Bates, Condon, Mroczek, and Revelle, 2018; Seeboth and Mõttus, 2018). To achieve a similar model, researchers could compute beta coefficients from the partial correlation coefficients for the outcome variable (see Haslbeck and Waldorp, 2018). More directly, researchers could square these same partial correlation coefficients to derive partial R^2 values (or the residual variation explained by adding the variable to the network), which makes for more interpretable results (Haslbeck and Waldorp, 2018).

Within our proposed framework, networks would be composed of personality components rather than specific items or facets. On the surface, this seems to clash with recent articles (including this special issue) demonstrating the unique predictive value of items in personality–outcome associations. In our view, this conflict is relatively minimal; instead, we think that unique personality components should be tapping into the very same notion. Items should have unique predictive value if they have distinct causes beyond other items—just as personality components are conceptualized.

On the one hand, when considering the items, 'Hate being the center of attention', 'Make myself the center of attention', 'Like to attract attention', and 'Dislike being the center of attention', there is unlikely to be unique predictive value of one item over another. On the other hand, unique items that do not have such an obvious overlap should remain as items and therefore unique components in the personality network. Therefore, we view personality components to be completely compatible with the unique predictive value of items while reducing homogeneous sets of items to their unique causes.

`Finally, researchers may be interested in the relations between traits and outcomes. As mentioned in our section on validity, traits are viewed as a summary of the network's state. Based on this definition, a summary statistic could be computed and used to evaluate the relationship between traits and outcomes. Using network loadings, Christensen and Golino (2020) proposed multiplying the loading matrix by the observed data to derive a weighted composite for each dimension (e.g. facets and traits) in a personality network. These composites could then be used in traditional analyses (e.g. zero-order correlation and regression) or as variables in a 'higher order' network with the outcome (and covariate) variables included. Following the same suggestions above, researchers could then square the outcome's partial correlation coefficients to derive the partial R^2 for the higher order personality components in the network.

CONCLUSION

So what are personality traits? At this point, it's clear that how researchers answer this question should affect the psychometric model they choose. In doing so, there are different considerations that should be made when developing and selecting items for their scales as well as how they

should interpret the measures used to quantify their scales. In this article, we take the initial steps towards how researchers can go about this with psychometric network models. We by no means suggest that our views represent the views of all researchers using these models (including latent variable models); however, we have provided a foundation for future work and discussion. Undoubtedly, the successful application of psychometric network models in personality psychology requires explicit definition and formalization of their measurement (e.g. components), which we have provided (Costantini and Perugini, 2012; Cramer et al., 2012b). We are optimistic that the continued development of measurement from a psychometric network perspective can move the theoretical and substantive assessment of personality traits forward.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting info item

REFERENCES

Afzali, M. H., St`ewart, S. H., Séguin, J. R., & Conrod, P. (2020). The network constellation of personality and substance use: Evolution from early to late adolescence. *European Journal of Personality*. https://doi.org/10.1002/per.2245

Allport, G. W. (1961). *Pattern and growth in personality*. Oxford, UK:Holt, Reinhart, & Winston.

Ashton, M. C., & Lee, K. (2005). A defence of the lexical approach to the study of personality structure. *European Journal of Personality*, 19, 5–24. https://doi.org/10.1002/per.541

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., Costantini, G., Denissen, J. J. A., Fleeson, W., Grafton, B., Jayawickreme, E., Kurzius, E., MacLeod, C., Miller, L. C., Read, S. J., Roberts, B., Robinson, M. D., Wood, D., & Wrzus, C. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality*, 31, 503–528. https://doi.org/10.1002/per.2115

Blanken, T. F., Deserno, M. K., Dalege, J., Borsboom, D., Blanken, P., Kerkhof, G. A., & Cramer, A. O. J. (2018). The role of stabilizing and communicating symptoms given overlapping communities in psychopathology networks. *Scientific Reports*, 8, 5854. https://doi.org/10.1038/s41598-018-24224-2

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time? *Journal of Research in Personality*, 32, 202–221. https://doi.org/10.1006/jrpe.1997.2206

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71, 425–440. https://doi.org/10.1007/s11336-006-1447-6

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, 64, 1089–1108. https://doi.org/10.1002/jclp.20503

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16, 5–13. https://doi.org/10.1002/wps.20375

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Franić, S. (2009). The end of construct validity. In Lissitz, R. W. (Ed.), *The concept of validity: Revisions, new directions and*

applications. Charlotte, NC:IAP Information Age Publishing, pp. 135–170.

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PLoS ONE*, *6*. https://doi.org/10.1371/journal.pone.0027407

Borsboom, D., & Mellenbergh, G. J. (2007). Test validity in cognitive assessment. In Leighton, J. P., & Gierl, M. J. (Eds.), *Cognitive diagnostic assessment for education: Theory and applications*. New York, NY:Cambridge University Press, pp. 85–116. https://doi.org/10.1017/cbo9780511611186.004

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological review*, *110*, 203–219. https://doi.org/10.1037/0033-295X.110.2.203

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., Wigman, J., & Snippe, E. (2019). What do centrality measures measure in psychology networks? *Journal of Abnormal Psychology*, *128*, 892–903. https://doi.org/10.1037/abn0000446

Bringmann, L. F., & Eronen, M. I. (2018). Don't blame the model: Reconsidering the network approach to psychopathology. *Psychological Review*, *125*, 606–615. https://doi.org/10.1037/rev0000108

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Cattell, R. B. (1946). *Description and measurement of personality*. New York, NY:World Book Company.

Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annual Review of Psychology*, *56*, 423–452. https://doi.org/10.1146/annurev.psych.56.091103.070133

Christensen, A. P. (2020). *Towards a network psychometrics approach to assessment: Simulations for redundancy, dimensionality, and loadings (Unpublished doctoral dissertation). University of North Carolina at Greensboro, Greensboro, NC, USA*. https://doi.org/10.31234/osf.io/84kgd

Christensen, A. P., Cotter, K. N., & Silvia, P. J. (2019). Reopening openness to experience: A network analysis of four openness to experience inventories. *Journal of Personality Assessment*, *101*, 574–588. https://doi.org/10.1080/00223891.2018.1467428

Christensen, A. P., & Golino, H. (2020). Statistical equivalency of factor and network loadings. *PsyArXiv*. https://doi.org/10.31234/osf.io/xakez

Christensen, A. P., & Golino, H. (2019). Estimating the stability of the number of factors via Bootstrap Exploratory Graph Analysis: A tutorial. *PsyArXiv*. https://doi.org/10.31234/osf.io/9deay

Condon, D. M. (2018). The SAPA personality inventory: An empirically-derived, hierarchically-organized self-report personality assessment model. *PsyArXiv*. https://doi.org/10.31234/osf.io/sc4p9

Connelly, B. S., Ones, D. S., Davies, S. E., & Birkland, A. (2014). Opening up openness: A theoretical sort following critical incidents methodology and a meta-analytic investigation of the trait family measures. *Journal of Personality Assessment*, *96*, 17–28. https://doi.org/10.1080/00223891.2013.809355

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. J. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, *54*, 13–29. https://doi.org/10.1016/j.jrp.2014.07.003

Costantini, G., & Perugini, M. (2012). The definition of components and the use of formal indexes are key steps for a successful application of network analysis in personality psychology. *European Journal of Personality*, *26*, 434–435. https://doi.org/10.1002/per.1869

Costantini, G., Richetin, J., Borsboom, D., Fried, E. I., Rhemtulla, M., & Perugini, M. (2015). Development of indirect measures of conscientiousness: Combining a facets approach and network

analysis. *European Journal of Personality*, *29*, 548–567. https://doi.org/10.1002/per.2014

Costantini, G., Richetin, J., Preti, E., Casini, E., Epskamp, S., & Perugini, M. (2019). Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences*, *136*, 68–78. https://doi.org/10.1016/j.paid.2017.06.011

Cramer, A. O. J. (2012). Why the item "23+ 1" is not in a depression questionnaire: Validity from a network perspective. *Measurement: Interdisciplinary Research & Perspective*, *10*, 50–54. https://doi.org/10.1080/15366367.2012.681973

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012a). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*, 414–431. https://doi.org/10.1002/per.1866

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012b). Measurable like temperature or mereological like flocking? On the nature of personality traits. *European Journal of Personality*, *26*, 451–459. https://doi.org/10.1002/per.1879

Cramer, A. O. J., Waldrop, L. J., van der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*, 137–150. https://doi.org/10.1017/S0140525X09991567

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334. https://doi.org/10.1007/BF02310555

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. https://doi.org/10.1037/h0040957

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*, 1–34. https://doi.org/10.18637/jss.v064.i04

Devellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Thousand Oaks, CA:SAGE Publications.

Deyoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*, 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Dinić, B. M., Wertag, A., Tomašević, A., & Sokolovska, V. (2020). Centrality and redundancy of the Dark Tetrad traits. *Personality and Individual Differences*, *155*, 109621. https://doi.org/10.1016/j.paid.2019.109621

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*, 399–412. https://doi.org/10.1111/bjop.12046

Epskamp, S. (2019). *Psychonetrics: Structural equation modeling and confirmatory network analysis. R package verson 0.3.3*. https://CRAN.R-project.org/package=psychonetrics

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*, 617–634. https://doi.org/10.1037/met0000167

Epskamp, S., Maris, G., Waldrop, L. J., & Borsboom, D. (2018a). Network psychometrics. In Irwing, P., Hughes, D., & Booth, T. (Eds.), *The Wiley handbook of psychometric testing, 2 volume set: A multidisciplinary reference on survey, scale and test development*. New York, NY:Wiley. https://doi.org/10.1002/9781118489772.ch30

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018b). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *4*, 1–28. https://doi.org/10.1080/00273171.2018.1454823

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*, 370–378. https://doi.org/10.1177/1948550617693063

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*, 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*, 1097–1114. https://doi.org/10.1037/a0016786

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality*, *56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009

Forbes, M. K., Wright, A. C., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, *126*, 969–988. https://doi.org/10.1037/abn0000276

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2019). Quantifying the reliability and replicability of psychopathology network characteristics. *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2019.1616526

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*, 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*, 432–441. https://doi.org/10.1093/biostatistics/kxm045

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*, 26–34. https://doi.org/10.1037/0003-066X.48.1.26

Golino, H., & Christensen, A. P. (2020). *EGAnet: Exploratory Graph Analysis: A framework for estimating the number of dimensions in multivariate data using network psychometrics*. https://CRAN.R-project.org/package=EGAnet

Golino, H., & Demetriou, A. (2017). Estimating the dimensionality of intelligence like data using Exploratory Graph Analysis. *Intelligence*, *62*, 54–70. https://doi.org/10.1016/j.intell.2017.02.007

Golino, H., & Epskamp, S. (2017). Exploratory Graph Analysis: A new approach for estimating the number of dimensions in psychological research. *PloS ONE*, *12*, e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of Exploratory Graph Analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*. https://doi.org/10.1037/met0000255

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, *37*, 827–838. https://doi.org/10.1177/001316447703700403

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, *18*, 277–296.

Hallquist, M., Wright, A. C. G., & Molenaar, P. C. M. (2019). Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2019.1640103

Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. M. (2007). The integrated trait–state model. *Journal of Research in Personality*, *41*, 295–315. https://doi.org/10.1016/j.jrp.2006.04.003

Haslbeck, J. M. B., & Waldorp, L. J. (2015). mgm: Structure estimation for time-varying mixed graphical models in high-dimensional data. *arXiv*. https://arxiv.org/abs/1510.06871

Haslbeck, J. M. B., & Waldorp, L. J. (2018). How well do network models predict observations? On the importance of predictability in network models. *Behavior Research Methods*, *50*, 853–861. https://doi.org/10.3758/s13428-017-0910-x

Hogan, R., & Foster, J. (2016). Rethinking personality. *International Journal of Personality Psychology*, *2*, 37–43. https://ijpp.rug.nl/article/view/25245

Hubley, A. M., Zhu, S. M., Sasaki, A., & Gadermann, A. M. (2014). Synthesis of validation practices in two assessment journals: Psychological Assessment and the European Journal of Psychological Assessment. In Zumbo, B., & Chan, E. (Eds.), *Validity and validation in social, behavioral, and health sciences*. Cham, CH:Springer, pp. 193–213.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives, (2nd ed.). In Pervin, L. A, & John, O. P. (Eds.), *Handbook of personality: Theory and research*. New York, NY:Guilford Press, pp. 159–181.

Kan, K.-J., van der Maas, H. L. J., & Levine, S. Z. (2019). Extending psychometric network analysis: Empirical evidence against g in favor of mutualism? *Intelligence*, *73*, 52–62. https://doi.org/10.1016/j.intell.2018.12.004

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*, 1–73. https://doi.org/10.1111/jedm.12000

Kelley, T. L. (1927). *Interpretation of educational measurements*. New York, NY:MacMillan.

Kruis, J., & Maris, G. (2016). Three representations of the Ising model. *Scientific Reports*, *6*, srep34175. https://doi.org/10.1038/srep34175

Lauritzen, S. L. (1996). *Graphical models*. Oxford, UK:Clarendon Press.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*, 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, *25*, 543–556. https://doi.org/10.1177/1073191116659134

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635–694.

Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York, NY: Routledge. https://doi.org/10.4324/9780203501207

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., van der Maas, H. L. J., & Maris, G. (2018). An introduction to network psychometrics: Relating Ising network models to item response theory models. *Multivariate Behavioral Research*, *53*, 15–35. https://doi.org/10.1080/00273171.2017.1379379

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, *19*, 97–112. https://doi.org/10.1177/1088868314541857

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81–90. https://doi.org/10.1037/0022-3514.52.1.81

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality, (3rd ed.). In John, O. P., Robins, R. W., & Pervin, L. A. (Eds.), *Handbook of personality: Theory and research*. New York, NY:Guilford Press, pp. 159–181.

McCrae, R. R., Costa, P. T., Ostendorf, F., Angleitner, A., Hřebíčková, M., Avia, M. D., Sanz, J., Sanchez-Bernardos, M. L., Kusdil, M. E., Woodfield, R., Saunders, P. R., & Smith, P. B. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, *78*, 173–186. https://doi.org/10.1037//0022-3514.7S.1.17

McCrae, R. R., & Mõttus, R. (2019). A new psychometrics: What personality scales measure, with implications for theory and assessment. *Current Directions in Psychological Science*.

McDonald, R. P. (1999). *Test theory: A unified treatment*. New York, NY:Taylor & Francis. https://doi.org/10.4324/9781410601087

McDonald, R. P. (2003). Behavior domains in theory and in practice. *Alberta Journal of Educational Research*, *49*, 212–230.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*, 412–433. https://doi.org/10.1037/met0000144

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*, 5–11. https://doi.org/10.3102/0013189X018002005

Messick, S. (1995). Standards of validity and the validity of standards in performance asessment. *Educational measurement: Issues and practice*, *14*, 5–8. https://doi.org/10.1111/j.1745-3992.1995.tb00881.x

Mõttus, R., & Allerhand, M. (2017). Why do traits come together? the underlying trait and network approaches. In Ziegler-Hill, V., & Shackelford, T. K. (Eds.), *SAGE handbook of personality and individual differences: The science of personality and individual differences*. London, UK:SAGE Publications, pp. 1–22.

Mõttus, R., Bates, T., Condon, D. M., Mroczek, D., & Revelle, W. (2018). Leveraging a more nuanced view of personality: Narrow characteristics predict and explain variance in life outcomes. *PsyArXiv*. https://doi.org/10.31234/osf.io/4q9gv

Nowick, K., Gernat, T., Almaas, E., & Stubbs, L. (2009). Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proceedings of the National Academy of Sciences*, *106*, 22358–22363. https://doi.org/10.1073/pnas.0911376106

Pérez, M.-E., & Pericchi, L. R. (2014). Changing statistical significance with the amount of information: The adaptive α significance level. *Statistics & Probability Letters*, *85*, 20–24. https://doi.org/10.1016/j.spl.2013.10.018

Perugini, M., Costantini, G., Hughes, S., & De Houwer, J. (2016). A functional perspective on personality. *International Journal of Psychology*, *51*, 33–39. https://doi.org/10.1002/ijop.12175

Pervin, L. A. (1994). A critical analysis of current trait theory. *Psychological Inquiry*, *5*, 103–113. https://doi.org/10.1207/s15327965pli0502_1

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, *10*, 191–218. https://doi.org/10.7155/jgaa.00185

R Core Team (2020). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rauthmann, J. F., Horstmann, K. T., & Sherman, R. A. (2019). Do self-reported traits and aggregated states capture the same thing? A nomological perspective on trait-state homomorphy. *Social Psychological and Personality Science*, *10*, 596–611. https://doi.org/10.1177/1948550618774772

Rauthmann, J. F., & Sherman, R. A. (2018). The description of situations: Towards replicable domains of psychological situation characteristics. *Journal of Personality and Social Psychology*, 482–488. https://doi.org/10.1037/pspp0000162

Read, S. J., Monroe, B. M., Brownstein, A. L., Yang, Y., Chopra, G., & Miller, L. C. (2010). A neural network model of the structure and dynamics of human personality. *Psychological Review*, *117*, 61–92. https://doi.org/10.1037/a0018131

Revelle, W. (2019). *psychTools: Tools to accompany the 'psych' package for psychological research*, Northwestern University, Evanston, Illinois. https://CRAN.R-project.org/package=psychTools

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, *45*, 73–103. https://doi.org/10.1080/00273170903504810

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*, 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D.(2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, *31*, 43–53. https://doi.org/10.1016/j.newideapsych.2011.02.007

Schwaba, T., Rhemtulla, M., Hopwood, C. J., & Bleidorn, W. (2020). The facet atlas: Using network analysis to describe the blends, cores, and peripheries of personality structure. *PsyArXiv*. https://doi.org/10.31234/osf.io/zskfu

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European Journal of Personality*, *32*, 186–201. https://doi.org/10.1002/per.2147

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*, 107–120. https://doi.org/10.1007/S11336-008-9101-0

Sočan, G. (2000). Assessment of reliability when test items are not essentially τ-equivalent. *Developments in Survey Methodology*, *15*, 23–35.

van Bork, R., Rhemtulla, M., Waldorp, L. J., Kruis, J., Rezvanifar, S., & Borsboom, D. (2019). Latent variable models and networks: Statistical equivalence and testability. *Multivariate Behavioral Research*, 1–24. https://doi.org/10.1080/00273171.2019.1672515

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861. https://doi.org/10.1037/0033-295X.113.4.842

Wood, D., Gardner, M. H., & Harms, P. D. (2015). How functionalist and process approaches to behavior can explain trait covariation. *Psychological Review*, *122*, 84–111. https://doi.org/10.1037/a0038423

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*, 17. https://doi.org/10.2202/1544-6115.1128

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ωh. *Applied Psychological Measurement*, *30*, 121–144. https://doi.org/10.1177/0146621605278814