Published in final edited form as:

Psychol Assess. 2009 September ; 21(3): 272–284. doi:10.1037/a0016699.

On the Value of Homogeneous Constructs for Construct Validation, Theory Testing, and the Description of Psychopathology

Gregory T. Smith,

Department of Psychology, University of Kentucky

Denis M. McCarthy, and

Department of Psychological Sciences, University of Missouri

Tamika C. B. Zapsolski

Department of Psychology, University of Kentucky

Abstract

The authors argue for a significant shift in how clinical psychology researchers conduct construct validation and theory validation tests. They argue that sound theory and validation tests can best be conducted on measures of unidimensional or homogeneous constructs. Hierarchical organizations of such constructs are useful descriptively and theoretically, but higher order composites do not refer to definable psychological processes. Application of this perspective to the approach of the *Diagnostic and Statistical Manual of Mental Disorders* to describing psychopathology calls into doubt the traditional use of the syndromal approach, in which single scores reflect the presence of multidimensional disorders. For many forms of psychological dysfunction, this approach does not appear optimal and may need to be discarded. The authors note that their perspective represents a straightforward application of existing psychometric theory, they demonstrate the practical value of adopting this perspective, and they provide evidence that this shift is already under way among clinical researchers. Description in terms of homogeneous dimensions provides improved validity, utility, and parsimony. In contrast, the use of composite diagnoses can retard scientific progress and hamper clinicians' efforts to understand and treat dysfunction.

Keywords

valid diagnosis; construct validation; construct definitions; homogeneity; diagnostic progress

This article has two basic aims. The first is to advance the argument that unidimensional construct measures provide the best basis for construct validation tests and theory tests. Validation tests on multidimensional measures, which are composites of related measures, can obscure important psychological processes. Elements of composites can, and do, act differently from one another, so analysis of composite scores combines the potentially different roles of its elements. For that reason, the use of composite scores can lead to unclear conclusions. The second aim is to apply this perspective to the problem of assessing and describing psychopathology. Many psychiatric diagnoses are composites of more than one construct, and so they may not represent meaningful psychological entities for most scientific purposes. Their use can interfere with clinical practice. We therefore advocate a new approach to

psychopathology description that is based on homogeneous dimensions of dysfunction. Below we discuss the implications of this approach for traditional, syndrome-based diagnosis.

We believe this position follows from a straightforward application of validity theory and psychometric theory. As we describe below, the departure is already well under way, and the approach we advocate requires no new tools for researchers. We argue that a focus on homogeneous constructs is both necessary and quite practical. At the same time, we appreciate that this proposal may represent a departure from much standard practice in clinical research and from past diagnostic systems. We therefore seek to articulate and discuss objections to our argument over the course of the article.

To make our argument, we begin by placing this discussion in the broader framework of scientific activity. We then explain our advocacy of the use of unidimensional constructs in theory testing and discuss its application in the domain of personality description. In that section of the article, we consider the different roles of lower order and higher order constructs. We then consider the implications of this perspective for the next generation of diagnostic systems with respect to both scientific inquiry and clinical application. In that section of the article, we discuss the implications of our perspective for traditional, syndromal approaches to psychological diagnosis.

The Progressive Refinement of Knowledge as Characteristic of Science

Scientific advancement is characterized, in part, by progressively more precise understandings of aspects of the physical and natural worlds, with a resulting increased capacity to explain, predict, and control events. From recognizing that the atom has distinct parts to identifying the presence of exons and introns within genes (parts used to make and not to make proteins, respectively), scientists have made progressive refinements in understanding that facilitate the explanation of what had previously been unexplainable. In virtually every area of science, increased precision in our understanding of phenomena is an active, ongoing enterprise.

Consider a few recent examples. First, in medical genetics there is a set of overlapping disorders known as Noonan syndrome, LEOPARD syndrome, neurofibromatosis Type 1, and neurofibromatosis–Noonan syndrome, and all tend to be characterized by facial dysmorphisms, short stature, and congenital heart defects (Sarkozy et al., 2007). However, recent molecular analysis of aspects of target genes led to the conclusion that the overlap between the former two disorders and the latter two was not paralleled by common molecular events (De Luca et al., 2005). That the two sets of disorders have distinct genetic causes has led to improved, more specific clinical recommendations (Sarkozy et al., 2007). Despite their morphological similarity, which resulted in their similar names, the two sets of disorders have different genetic etiologies. Entities once thought to be the same were recognized not to be.

Second, in the clinical realm a common task is to recognize distinctions among similar phenomena that require different forms of intervention (Fukayama & Osawa, 2006). In 1969, the number of different recognized epileptic syndromes was 125; today, there are over 300, and the new distinctions have helped identify different gene defects in different patients and have led to different treatments (Fukayama & Osawa, 2006). Similarly, the progressive recognition of meaningful heterogeneity among breast cancers and gynecological cancers has improved treatment success for each (Bunnell & Winer, 2002; Greven, 2005). Chronic pain treatments have been improved as researchers have come to recognize important distinctions among patients (Turk, 2005). Across diverse areas of science, a continual process of identifying new distinctions among entities takes place and bears clinical fruit (Morgan, 1997).

A second aspect of scientific advancement is the integration of lower order phenomena into higher order, integrative theories. In many areas of science, the pursuit of integrative theories

is a central task of scientists. The pursuit of a unified field theory in physics (Georgi & Glashow, 1972); the development of theories to integrate biological evolutionary processes at the multiple levels of gene, organism, and species (Keller, 1999); and, closer to home, the development of comprehensive theories of human personality (Costa & McCrae, 1994; Digman, 1990) are among the many examples of this process. Integrative theory is fundamentally important: It organizes diverse phenomena into meaningful wholes; it provides insight into etiology and, in clinical science, intervention; and it generates hypotheses that lead to the identification of previously unrecognized processes.

The two types of advances—that is, increased differentiation of lower order phenomena and integration of those phenomena into larger, explanatory theories—are closely linked. In particular, the validity of integrative theories depends on the validity of the elements contributing to those theories, just as identification of important, lower order elements is contributed to by sound integrative theories. One set of implications of this perspective for clinical psychological science includes the following: When it occurs that a previously recognized psychological construct is subdivided into more elemental components that have different etiologies, or different external correlates, or that require different interventions, it no longer makes sense to treat the original entity as a coherent, homogeneous construct and to represent it by a single score. If one were to refer to such an entity with a single score, one would risk errors at both the lower order level and the higher order level. At the lower order level, one would risk compromising accurate description and effective intervention. At the higher order level, one would risk compromising valid aggregation into larger theories. To clarify the basis for this claim, we next discuss validity testing of clinically relevant constructs and the centrality of unidimensional constructs in this enterprise.

Construct Validity/Theory Testing: The Importance of Unidimensionality

Clinical psychology theory involves specifying the nature of relationships among different psychological entities (causal, correlate, mediator, moderator, and so on). One of the fascinating challenges of psychological science is that the psychological entities we study cannot be directly observed (Cronbach & Meehl, 1955); researchers must infer their existence. Researchers do so in order to best approximate their understanding of real psychological phenomena (Borsboom, Mellenbergh, & van Heerden, 2004). Their doing so has clear utility for helping us understand human behavior, differences among individuals, and dysfunction (Smith, 2005).

Psychologists test their theories by developing measures of the inferred entities and testing whether the measures relate to measures of other inferred entities as specified by theory. Repeated findings consistent with theory produce increasing confidence in both the theory and the measures used to represent the constructs of interest. Findings inconsistent with the theory likewise raise questions about both the theory and the measures used to test it. The indeterminate nature of this process is clear and has been described many times before (see Smith, 2005; Strauss & Smith, 2009).

It follows that when we refer to construct validity studies, we are necessarily referring to simultaneous tests of psychological theories and of psychological measures (Cronbach & Meehl, 1955; Smith, 2005). The process of construct validation requires clear, coherent definitions of target constructs and a clear statement of anticipated relationships between the target construct and other constructs. Tests of the validity of construct measures must inevitably be tests of theories specifying relationships among the constructs. A key point for the present discussion is this: To the degree that one uses a single score from a target measure that includes multiple dimensions (such as a measure of posttraumatic stress disorder thought to include four factors, or a measure of extraversion thought to have six facets), one's construct validation/

theory test has theoretical uncertainty built in. Such a test is likely to have reduced scientific value.

If one correlates a total score of a multidimensional measure with a criterion, one builds two sources of uncertainty into one's test. One source of uncertainty is that, with a single score, one cannot know the nature of the different dimensions' contribution to that score. Conceivably, an overall correlation could reflect the same magnitude of relation between each dimension and the criterion, but that may well not be true. It is more likely that such a correlation reflects a kind of average of strong and weak relationships between different dimensions and the criterion (Smith, Fischer, & Fister, 2003; Smith & McCarthy, 1995). One cannot know the meaning of a single score representing a multidimensional measure (Borsboom et al., 2004; McGrath, 2005).

There is an extensive history of arguments made by psychometricians that underlie this position. Edwards (2001) noted that researchers have long appreciated the need to avoid heterogeneous items: If such an item predicts a criterion, one will not know which aspect of the item accounts for the covariance. The same reasoning extends to tests: If a test includes multiple dimensions, one cannot know which dimensions account for the test's covariance with measures of other constructs. If one uses single scores from multidimensional tests, one has simply moved the heterogeneity problem from the item level to the scale level (Smith et al., 2003). Hough and Schneider (1995); McGrath (2005); Paunonen and Ashton (2001); and Schneider, Hough, and Dunnette (1996), among others, have all noted that use of scores of broad measures often obscures predictive relationships. Indeed, studies comparing prediction using specific facets of broad personality dimensions with prediction using scores on the dimensions themselves show that prediction is improved when one represents each facet individually (Paunonen, 1998; Paunonen & Ashton, 2001). Essentially, one gives oneself the chance to study the separate and incremental roles of each dimension involved in one's measures, rather than averaging across the different dimensions before predicting.

It is not just that a composite score averages the functioning of separate constructs in its association with measures of other constructs. The problem is more severe than that. A second source of uncertainty is that the same composite score will tend to reflect different combinations of construct scores for different individuals in a sample. For example, imagine two individuals with the same overall Neuroticism score on the Revised NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992) measure of the five-factor model (FFM) of personality. Two of the six facets of Neuroticism in that measure are angry hostility and anxiety. One person could be high in angry hostility but low in anxiety, and the other could be low in angry hostility but high in anxiety. This possibility is not just hypothetical: In the standardization sample, the two traits correlated $r = .47$, meaning they shared only 22% of their variance (Costa & McCrae, 1992). Thus, covariation of an overall Neuroticism score with another variable lacks clear meaning: The Neuroticism score likely reflects different patterns of traits for different individuals. It is not just that when one correlates neuroticism with another variable one cannot know whether the correlation was "carried" by, in this case, angry hostility or anxiety; it is that the same score could reflect angry hostility elevations for some individuals and anxiety elevations for others. For these reasons, the central construct validation process should be to test hypothesized relationships among what are thought to be homogeneous, precisely defined constructs.

Issues Concerning the Determination of Homogeneity

As compelling as the above example may be, an obvious question concerns how one determines homogeneity in the first place. In our view, the statement that a construct is homogeneous is a statement of theory that, itself, needs to be investigated empirically. Just as theory validation is an indeterminate and ongoing process, so is determination of homogeneity. Suppose that

one's theory holds that the construct is a single dimension, factor analyses have supported the single dimension hypothesis, and there is no evidence that different components of the construct play different roles in understanding psychological processes (e.g., different components are not understood to have different correlates with other constructs). When that is the case, researchers can and should appropriately treat their measure of the construct as unidimensional and test theories on that basis.

This outcome does not preclude the possibility that at some future point, heterogeneity within the construct may be uncovered. When that occurs, the new understanding of what is unidimensional should guide theory testing efforts. Ongoing investigations of the structure of constructs, conducted with different comparison measures and different samples, provide new information concerning claims of homogeneity. Part of the indeterminate nature of theory testing is the indeterminate nature of the validity of constituent constructs, which includes the structure of those constructs.

It is important to appreciate that determination of homogeneity, like the determination of validity (Borsboom et al., 2004), must be based on theory and cannot be answered solely through statistics. Empirical evidence concerning dimensionality should be evaluated with respect to the theoretical appropriateness of the test. For example, one cannot view emergence of a single factor in factor analysis as necessarily indicative of unidimensionality. If one were to conduct a factor analysis on items representing a broad range of constructs, one whole domain may fall on a single factor and thus seem unidimensional. But a factor analysis on items restricted to that domain may identify subdomains within it, and that result suggests that the domain is multidimensional. The latter test may be a more appropriate test of one's theory of dimensionality.

A related concern involves the risk of infinite reduction; after all, individual items differ from each other and do not covary perfectly. Does each item represent its own dimension? When can we conclude we are measuring constructs at the level of homogeneity?

Suppose one believes a measure actually represents two dimensions and thus subsumes two constructs. Because determination of dimensionality is a theory testing process, there are at least two pieces of relevant empirical evidence. First, one is likely to conduct a structural test of dimensionality, such as factor analysis or latent class analysis, on theoretically appropriate samples. Second, to say that there are two dimensions is to imply that the two dimensions play different roles in psychological theory. If two putative dimensions do not play different roles in theory—that is, if, in every case, measures of the two dimensions correlate the same with measures of other constructs; predict the same external behaviors, attitudes, and cognitions; are equally heritable; and are related to the same gene polymorphisms—then there is no evidence that the two measures reflect meaningfully different psychological processes. The use of two terms and two measures would be both unnecessary and potentially misleading. The validity of the claim that the two are distinct has been compromised.

Our position that determinations of dimensionality are, in part, a function of whether putative dimensions play different roles in relation to other psychological processes is based on the usefulness of distinctions for psychological theory and clinical application. When two constructs have been shown to be distinct, in the sense that they appear to play different roles in relation to other psychological processes, then researchers and clinicians will typically benefit by distinguishing between them. When they have not been shown to be distinct in this way, then there is neither a practical benefit of doing so nor a theoretical necessity to do so. As William James put it, “There can be no difference anywhere that doesn't make a difference elsewhere—no difference in abstract truth that doesn't express itself in a difference in concrete fact” (as cited in McDermott, 1967, p. 379).

Thus, determinations of homogeneity are tests of theories that are thus always subject to revision as knowledge develops, and claims of dimensionality require a demonstration of the different roles of the different putative dimensions in psychological theory and explanation. Of course, the reality that determinations of homogeneity are, in fact, indeterminate is not a reason to disregard known heterogeneity in construct measures.

Hierarchical Organizations of Homogeneous Measures

Developing measures of cohesive, homogeneous entities is crucial to psychological assessment science. As researchers do so, they face the related challenge of how to organize such constructs into informative, descriptive theoretical systems, such as hierarchies. Hierarchies and other organizational structures can enhance understanding by providing description across varying levels of abstraction (Digman, 1997; Markon, Krueger, & Watson, 2005; Morgan, 1997). Of course, such systems depend, for their accuracy, on the validity of their elemental components.

It is important to appreciate what hierarchical or other organizational systems provide and what they do not provide. They provide potentially useful theoretical, descriptive accounts of the relations among homogeneous constructs. Organizational frameworks of constructs—such as the five-, four-, three-, and two-factor models of personality (Markon et al., 2005)—provide a valuable sense of which constructs tend to covary more with each other than with other constructs and hence enhance understanding of psychological processes.

However, higher order factors, developed as summaries or other mathematical combinations of homogeneous constructs, cannot refer to single, meaningful psychological entities and hence cannot refer to causally active constructs (McGrath, 2005). An overall score on Neuroticism does not provide information about the specific psychological processes in place. As Saucier (1998) put it, “Broad factors have the disadvantage of being in effect composed of many variables, and thus possessing some definitional ambiguity” (p. 264).

This claim may not appear to give higher order constructs their appropriate due. As L. A. Clark put it, “If you are advocating that we get rid of Neuroticism, that won't work. Neuroticism is both widely accepted and has been extremely useful” (personal communication, July 1, 2008). We agree. We are not advocating that we remove Neuroticism from the personality psychopathology discussion at all. Rather, we are advocating what we believe to be a more theoretically sound understanding of what Neuroticism represents. Using, for this purpose, the NEO PI-R five-factor model, in which Neuroticism is a higher order factor and there are six provisionally homogeneous traits within that umbrella (depression, anxiety, vulnerability, angry hostility, impulsiveness, and self-consciousness), we understand Neuroticism as denoting the finding that those six traits share substantial variance and have more in common than they have in common with other personality traits.

The recognition that those traits share substantial variance is of great importance for theory and understanding. The characterization of human personality as varying across five broad dimensions has clarified understanding, provided a sound basis for the evaluation of new claims concerning personality functioning, and integrated research across numerous domains. As one more specific example of the importance of Neuroticism, the six traits within Neuroticism appear to have both common and unique heritability: Some of their shared variance has a shared genetic basis (Jang, Livesley, & Vernon, 1996). Recognition of shared genetic etiology across traits suggests some commonality of cause for the different traits. When extended to psychopathology, findings of this kind provide an impetus for the development of models identifying underlying dimensions of dysfunction (Krueger & Markon, 2006). For this and other reasons, Neuroticism (and other broad traits) will continue to play a central, integrating role in our understanding of human personality.

At the same time, we advocate against using a single score to reflect variation on Neuroticism. We make this claim for the straightforward psychometric reasons described above: A common Neuroticism score can quite plausibly be obtained by individuals with different patterns of traits within the subjective distress domain. When Neuroticism is calculated as the sum of scores reflecting six different traits, then that sum lacks clear psychological meaning.

One potential alternative view is that the focus should not be on lower level, homogeneous constructs; rather, the real question concerns the optimal bandwidth, or optimal resolution, of measurement for a given purpose. Perhaps higher order constructs such as neuroticism are preferred for some purposes, and lower order constructs such as anxiety or angry hostility are preferred for others. We very much agree with an emphasis on the different roles of broad, multidimensional domains and precisely definable, homogeneous constructs. In fact, we believe our framework offers more precision to the discussion of the different roles of higher order and lower order constructs. There is not a sound psychometric basis for using single scores to reflect broad dimensions in theory tests or prediction studies, so homogeneous constructs should be used for such purposes. Broad construct domains such as neuroticism can provide a powerful, integrative perspective that helps clarify the roles of specific constructs. But if one wants to represent a broad domain in a theory test, one should include separate scores representing each specific construct subsumed in that domain; doing so enables one to know whether different specific constructs play different roles from each other.

Factor Analysis and the Identification of Homogeneity

As referred to above, factor analysis is a statistical tool that can help determine dimensionality when it is used to test theory. In this section, we briefly consider an important, but not always recognized, distinction between two different types of factors that also pertain to the determination of dimensionality. In one case, the indicators of a factor are a set of items determined to represent the same construct. The items or indicators are not understood to represent different constructs from each other; each item is instead understood to be an expression of the construct represented by a factor. When that is true, the factor represents a definable, homogeneous construct (provisionally, of course, pending ongoing evaluation of the validity of the theory and the empirical research). Two individuals with the same score on the factor would be understood to have the same level of the construct. This case is often described in terms of latent variable theory (Bollen, 2002; Bollen & Lennox, 1991; Borsboom, Mellenbergh, & van Heerden, 2003). Although a complete review of latent variable models is beyond the scope of this article (see Borsboom et al., 2003), the theoretical position is that variation in the latent variable causes variation in its indicators; each indicator is an expression of the latent variable.

In the second case, the items are not alternative expressions of the same underlying construct but are understood to represent different constructs that share variance with each other. One conducts this type of factor analysis in order to identify dimensions that describe which constructs tend to covary most highly. In this case, the factor does not represent a single, definable construct; it represents variance shared among a set of constructs. There is no reason to think that two individuals with the same score on this kind of factor are the same on the factor's constituent constructs.

Confirmatory factor analysis (CFA) and exploratory factor analysis (EFA), as they are typically used, test different models of the relations among variables. CFA tests the latent variable model. A typical CFA is structured such that each hypothesized factor is represented by multiple, corresponding indicators. When there are multiple latent variables in a model (i.e., multiple factors), typical CFA models are testing the view that each factor's indicators are expressions of that factor and of no other factor; loadings from the latent variable to indicators of other

variables are fixed to zero. Indicators of that factor are specified not to correlate with other factors, and the degree to which that specification accords with the data is evaluated with model fit indexes. Thus, the model to be tested specifies that each latent variable is homogeneous.

EFA does not test the latent variable model; it simply identifies dimensions of shared variance among the items factor-analyzed. That is because EFA does not impose the same specificity on a model. EFA-derived factors represent variance shared by a set of variables, but variables can and do load on more than one factor. Of course, simple structure is the typical goal for EFA analyses, but simple structure is neither absolute nor evaluated quantitatively. There is no single definition of adequate simple structure. Multiple factors can, and often do, reflect variance shared with the same variable. CFA, as typically used, imposes absolute simple structure.

The distinction between the two is not a trivial, inconsequential statistical fine point. It has real theoretical and practical importance. Each variable that loads on a factor derived via EFA cannot be understood to be an expression of a common, underlying factor, because variables typically have some additional loading on another factor. In contrast, variables that load on a single CFA factor can provisionally be understood to be alternative expressions of the same factor.

To illustrate this distinction, we continue our focus on personality theory, because current personality theory is characterized by well-developed theoretical models of hierarchy and extensive empirical description across multiple hierarchical levels. We believe the organizational structure within personality theory provides an example for other domains of clinical inquiry. As we show below, it is also true that applications of personality theory help clarify the meaning and dimensionality of psychopathological constructs.

Personality Structure

Personality has been studied extensively from a hierarchical perspective. It has been described as existing along two higher order dimensions known as alpha and beta (Digman, 1997); along three, four, or five dimensions that underlie those two (the five dimensions have labels such as neuroticism, extraversion, agreeableness, conscientiousness, and openness to experience; Costa & McCrae, 1992; Goldberg, 1993; Markon et al., 2005); and along 30 dimensions that underlie those five, higher order composites (Costa & McCrae, 1992, 1995). Here we focus primarily on studies of the NEO PI-R measure of the FFM of personality, both because there is extensive evidence of the validity of this measure of the FFM and because results with this measure represent the typical outcomes to comprehensive models of personality (Markon et al., 2005; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). The NEO PI-R representation of the FFM specifies six homogeneous facet constructs within each of the five broad factors, for a total of 30 personality constructs (Costa & McCrae, 1992). Figure 1 is a depiction of a common hierarchical personality model. It includes Digman's (1997) overarching two factors, the five factors as reflected in the NEO PI-R (Costa & McCrae, 1992), and the 30 specific, homogeneous NEO PI-R scales (e.g., Self-Consciousness, Excitement Seeking) at the lowest level.

EFAs of the FFM using the NEO PI-R across age, gender, race, nationality, and self versus observer report have repeatedly produced the five-factor structure in which the factors are labeled as described above (McCrae et al., 1996). The stability of the five-factor structure is striking: The five dimensions appear to summarize personality variability in a remarkably consistent way. However, as McCrae et al. (1996) noted, the FFM does not propose a simple structure, because many traits fall between the axes with the five-factor labels. Accordingly, one does get large secondary loadings for many of the 30 facets on the NEO PI-R. In one analysis reported by McCrae et al. (1996), the impulsiveness facet of the NEO PI-R neuroticism

dimension loaded .54 on neuroticism, $-.36$ on conscientiousness, .30 on extraversion, and $-.23$ on agreeableness. Each of those four broad factors shares variance with the impulsiveness scale. Angry hostility, another facet of neuroticism, typically has a large secondary negative loading on agreeableness. These multiple loadings make sense: The trait angry hostility conceptually and empirically falls in both the neuroticism and agreeableness domains. The traits measured by the facet scales are not meant to be, and indeed are not, simple expressions of the dimension with which they are primarily associated.

Consider the meaning of CFA tests of the FFM from this perspective. A CFA model specifies that each of the facets within a broad domain (such as neuroticism) is not in fact a separate construct; rather, each represents an alternative expression of the same factor. Thus, anxiety and angry hostility would be specified as alternative indicators of neuroticism, not as separate entities. And each of those traits would be specified as having zero loadings on the other four broad dimensions. The concept behind this modeling strategy is that the five factors each represent unidimensional psychological entities, and that the six facets for each factor are alternative manifestations of those underlying entities.

But this concept does not represent the existing theoretical understanding of the FFM (McCrae et al., 1996). Consistent with this discrepancy between the FFM and this CFA representation, CFA tests of the FFM with the NEO PI-R do not fit the data well (Church & Burke, 1994; McCrae et al., 1996). Whereas confirmatory fit index values of .90 or greater are considered indicative of acceptable model fit (Kline, 2005), McCrae et al. (1996) reported confirmatory fit index values of .55 for a model in which the five factors were orthogonal and .60 when the five factors were allowed to correlate. The poor fit is to be expected: It is not the case that the five factors define coherent, theoretically primary entities, such that variation in the five factors uniquely causes variation in their assigned facets. The facets were not designed to be, and are not understood as, alternative expressions of a common trait. That model is not the one described by McCrae et al. (1996), and it does not fit the data.

The conclusion is clear. The five factors of the FFM, as derived from EFA of the NEO PI-R, represent dimensions of variance shared by the 30 facet scales. They are not latent variables that “underlie” variation on the facet scales (and results are similar for other comprehensive models; Church & Burke, 1994; Markon et al., 2005). The five factors of the FFM provide useful, integrative knowledge about the covariance of personality traits, just as do other comprehensive models. But the factors cannot represent cohesive, homogeneous, theoretically active psychological entities.

This conclusion is neither new nor radical; it is entirely consistent with current FFM theory. McCrae et al. (1996) said that the FFM “does not assume that all personality traits define one and only one factor” (p. 553). They went on to say, “There is no theoretical reason why traits should not have meaningful loadings on three, four, or five factors” (p. 553). Costa and McCrae (1995) also noted that the five domains are not themselves mutually exclusive: Traits appear to relate to two or more of the five broad domains. In fact, they said they assigned facets to one and only one domain to accommodate the need for simplicity (Costa & McCrae, 1995). Hofstee, de Raad, and Goldberg (1992) also identified multiple loadings among traits. Authors have, consistent with this view, referred to the five domains as abstractions (Markon et al., 2005; Saucier, 1998). It seems clear that the traits measured by facet scales are not indicators in the CFA sense; they are not meant to be straightforward expressions of theoretically meaningful constructs (the five factors).

Implications of Factor Analyses of Personality for Validation and Theory Testing

There are clear implications of this analysis for construct and theory validation. One cannot explain an individual's perceptions or behavior by describing him or her as high on neuroticism.

The term lacks specific meaning. Among individuals with the same score on a measure of neuroticism, one person might be self-conscious and vulnerable but not particularly impulsive or angry. Another might be high in impulsivity and angry hostility, but feel neither self-conscious nor vulnerable. A high score on neuroticism does not describe a specific personality, so neuroticism itself cannot be understood to play a causal role in psychological theory.

The same is true of even higher order, more abstract combinations of traits. Digman's (1997) alpha, which he felt tended to reflect successful socialization, must be understood as a summary of separable, though related, processes. High scores on alpha can be obtained in many different ways (contributors to high scores could include measures of traits such as neuroticism, psychoticism, cognitive distortion, identity disturbance, affective instability, narcissism, alienation, and many others), so variation on alpha does not describe a psychological process and therefore cannot describe a causally active process. Only measures of unidimensional constructs can lay claim to explaining psychological processes and hence to explaining possible causal activity.

The Practicality of Focusing on Unidimensional Constructs

We have argued that a focus on unidimensional, homogeneous construct measures facilitates accurate theory and construct validation tests. We next provide examples to illustrate that doing so provides significant practical dividends for the advancement of assessment knowledge.

Consider the occupational variable service orientation to consumers (Hogan & Hogan, 1992). Although it might be appealing to hypothesize that conscientiousness, one of the five broad domains measured by the NEO PI-R, relates to service orientation, that hypothesis is imprecise. In fact, Costa and McCrae (1995) found that one trait within the conscientiousness domain, dutifulness, correlated .35 with service orientation, while another, achievement striving, correlated $-.01$ with the same criterion. Achievement striving did correlate highly with a different occupational variable, managerial potential ($r = .63$), but another facet of conscientiousness, order, accounted for comparatively little variance in that criterion ($r = .25$). We note two things about this example. First, these findings are typical: Facets within domains often have markedly different external correlates. Second, the findings do appear to support the validity of the facet scales; they make sense given the nature of the constructs being measured. Therefore, facet-level findings such as these lead to advances in understanding of psychological processes. They would not have occurred if Costa and McCrae (1995) had correlated only broad conscientiousness with the occupational variables. The precision gained by studying homogeneous constructs is fruitful.

A second example, not based on the NEO PI-R, concerns the construct impulsivity. Over the last several years, clinical researchers have recognized that the term *impulsivity* has been used in a variety of ways. Different measures with that label sometimes measure what appear to be different constructs, and many impulsivity measures appear to include items tapping multiple constructs (Bagby, Joffe, Parker, & Schuller, 1993; Depue & Collins, 1999; Evenden, 1999; Fischer, Smith, & Cyders, 2008; Petry, 2001; Smith et al., 2007; Whiteside & Lynam, 2001, 2003; Whiteside, Lynam, Miller, & Reynolds, 2005; Zuckerman, 1994).

Efforts to disaggregate the set of constructs included within the impulsivity framework have proven quite useful. One current model involves identification of five separate constructs that involve dispositions to rash action: sensation seeking, lack of planning, lack of perseverance, positive urgency (the tendency to engage in rash actions when in an extremely positive mood), and negative urgency (the tendency to engage in rash actions when in an extremely negative mood; Cyders & Smith, 2007, 2008b; Cyders et al., 2007; Smith et al., 2007; Whiteside & Lynam, 2001, 2003; Whiteside et al., 2005).

Measures of the five traits are only modestly correlated; the traits do not load on an overall “impulsivity” factor; and the traits have different external correlates (Smith et al., 2007). The urgency traits appear to relate to problem levels of involvement in risky behaviors, sensation seeking appears to relate to the frequency of engaging in risky behaviors, lack of planning relates to some problem behaviors but not others, and lack of perseverance relates to school performance (Cyders, Flory, Rainer, & Smith, 2009; Cyders & Smith, 2008a, 2008b; Fischer & Smith, 2008; Fischer, Smith, Annus, & Hendricks, 2007; Fischer et al., 2008; Miller, Flory, Lynam, & Leukefeld, 2003; Smith et al., 2007; Whiteside & Lynam, 2003; Whiteside et al., 2005). These different patterns of correlates are consistent with theory (Fischer, Smith, Spillane, & Cyders, 2005). In addition, different interventions are likely to be effective for different ones of the traits (perhaps distress tolerance for the urgency traits and safe, alternative ways to seek sensations for sensation seeking; Cyders & Smith, 2008b; Fischer & Smith, 2008; Palmgreen & Donohew, 2003). These advances could not have occurred, had researchers continued to rely on a single score from what turned out to be multidimensional measures of “impulsivity.”

These are two of many possible examples of the practical value of focusing on homogeneous, elemental personality constructs. Doing so is becoming increasingly common, is well within the expertise of clinical researchers, and produces meaningful advances in understanding.

Having laid this groundwork, we move next to the second purpose of this article and the focus of this special section, by applying our perspective to the assessment and description of psychopathology. As we show, in numerous domains researchers have demonstrated that putative “disorders” actually consist of multiple, homogeneous dimensions of dysfunction that have different correlates and, often, different etiologies. When this is true, the validity of the disorders as anything beyond a conventional abstraction is compromised: A diagnosis that a disorder is present, or an assignment of a symptom count for a disorder, can be misleading. We argue that diagnoses based on homogeneous dimensions of dysfunction have increased validity, and we further argue that such diagnoses can provide improved theoretical clarity, parsimony, and utility.

Psychopathology Description and Diagnosis

The approach of the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* to describing psychopathology makes use of the syndrome perspective, which involves identification of a constellation of symptoms thought to stem from a common cause or thought to indicate a disease or abnormal condition (Kraepelin, 1981). Disorders can include heterogeneous symptoms, as long as the various symptoms reflect a common cause. Thus, a valid syndrome is understood to reflect a homogeneous grouping of individuals (Robins & Guze, 1970). From a syndromal perspective, symptoms within a disorder may not correlate perfectly, even though they sometimes stem from a common cause and indicate a disease process. For example, the experience of headaches, muscle aches, sore throats, and fatigue do not covary perfectly because they can have different causes, but sometimes they occur together due to a common cause and are called the flu.

However, there is little evidence that most *DSM* disorders represent syndromes in this classic sense. As noted by Kupfer, First, and Regier (2002; Kupfer is chair and Regier vice chair of the *DSM-V* Task Force):

... the goal of validating these syndromes and discovering common etiologies has remained elusive. Despite many proposed candidates, not one laboratory marker has been found to be specific in identifying any of the *DSM*-defined syndromes. Epidemiologic and clinical studies have shown extremely high rates of comorbidities

among the disorders, undermining the hypothesis that the syndromes represent distinct etiologies. (p. xviii)

In some cases, in fact, there is evidence that the multiple dimensions within a disorder do have different etiologies. Many currently defined disorders appear to represent composites of multiple, separable constructs. Thus, the assignment of a disorder, use of disorder scores, or use of symptom counts as descriptions of clinical functioning or as variables in an analysis is problematic for at least two reasons. First, the scores represent the influence of multiple psychological constructs, so they lack clear theoretical meaning. Second, different individuals are likely attaining the same score through endorsement of different symptoms, so the relative degree of influence of the different constructs varies from person to person (McGrath, 2005). In such cases, the meaning of scores is not clear: The same diagnosis can be assigned to individuals experiencing meaningfully different forms of dysfunction.

Clinical researchers have long recognized the potential benefit of testing hypotheses at the symptom rather than syndrome level. For example, Persons (1986) advanced similar arguments for the importance of studying schizophrenia symptoms (hallucinations, delusions, disorganization, negative symptoms) rather than the potentially ambiguous classification of schizophrenia. Although her arguments were not based primarily on psychometric concerns, she argued that a focus on the symptom level reduces misclassification, enables researchers to study the specific phenomena of interest, and thus improves theoretical understanding of the various processes underlying the different schizophrenia symptoms.

Clinical researchers have also raised related concerns about disorder coherence. Two individuals could both be diagnosed with obsessive-compulsive personality disorder (OCPD) and yet not share a single symptom in common (Widiger & Trull, 2007), and there are over 100 different ways to meet the criteria for diagnosis of borderline personality disorder (Frances, First, & Pincus, 1995). Recognition of this problem has led clinical researchers to disaggregate multidimensional syndromes into homogeneous constructs. We next turn to examples of this work.

The Disaggregation of Mental Disorders

Psychopathy—The disaggregation of the many components of psychopathy has received considerable research attention (Brinkley, Newman, Widiger, & Lynam, 2004; Cooke & Michie, 2001; Harpur, Hakistan, & Hare, 1988; Harpur, Hare, & Hakistan, 1989; Lynam & Widiger, 2007). Hare's (2003) Psychopathy Checklist Revised (PCL-R) importantly identified two separate factors, one representing the callous and remorseless use of others and the other representing a deviant and antisocial lifestyle. In the PCL-R, the two factors share only 25% of their variance (Harpur et al., 1988), and they have numerous different correlates (Harpur et al., 1989). Cooke and Michie (2001) identified three factors, described as (a) arrogant and deceitful interpersonal style, (b) deficient affective experience, and (c) impulsive and irresponsible behavioral style. To complicate matters further, the PCL-R does not include all of the dimensions of the classic description of psychopathy provided by Cleckley (1941); for example, low anxiousness is not represented (Lynam & Widiger, 2007; Rogers, 1995). Brinkley et al. (2004) elaborated by arguing that psychopathy, as measured by the PCL-R, is an etiologically heterogeneous entity. If psychopathy includes multiple dimensions that do not always covary, and if those dimensions have different etiologies, then psychopathy may not be a coherent, meaningful psychological construct.

Most recently, Lynam and Widiger (2007) took advantage of the hierarchical disaggregation of personality to develop a comprehensive description of the psychopathy construct. For each of the 30 facets of the NEO PI-R, they identified whether (a) the trait related to psychopathy and (b) whether high or low trait scores reflected the psychopathy construct. The result is a

placement of psychopathy along each of 30 homogeneous dimensions of personality; in this view, psychopathy is understood to represent a multidimensional combination of constructs, rather than a coherent theoretical entity in and of itself. Lynam and Widiger's findings revealed meaningful distinctions between personality facets on the same broad personality domain, such as we described with respect to neuroticism.

Depression—Jang, Livesley, Taylor, Stein, and Moon (2004) studied the factor structure of depression. Using several symptom lists, they identified 14 subfactors. Examples of subfactors included “feeling blue and lonely,” “insomnia,” “positive affect,” “loss of appetite,” and “psychomotor retardation.” Interestingly, intercorrelations among the factors ranged from .00 to .34, and the factors were differentially heritable, with heritability coefficients ranging from .00 to .35. It appears to be the case that (a) some of the dimensions of depression do not covary substantially and (b) some have a heritable basis and others do not, which likely indicates that their etiologies differ. McGrath (2005) provided interesting examples of the heterogeneity of depression symptom items.

Perhaps, then, it is the case that depression is not a coherent, homogeneous psychological construct. It may instead be a hierarchical construct that involves shared variance among several, separable constructs (McGrath, 2005). Use of overall depression scores as a criterion in construct validity/theory testing studies is likely to be problematic. For example, to test whether stressful events are a risk factor for depression is imprecise: Are they a risk factor for each construct subsumed within the overall label? Are they a risk factor for only one construct, or for some subset of constructs? For example, do they tend to reduce positive affect but not influence negative affect? Or, do they increase negative affect but not relate to positive affect? Do they influence both? The imprecise test yields imprecise results. Following such a test, one does not have a coherent psychological finding.

Obsessive-compulsive disorder—Many authors have separated obsessive-compulsive disorder (OCD) into several dimensions. Watson and Wu (2005) identified obsessive checking, obsessive cleanliness, and compulsive rituals as separate and only moderately related constructs and concluded that OCD may be both phenotypically and genotypically heterogeneous. Leckman et al. (1997) found four dimensions within the OCD criteria that were intercorrelated between .50 and .56, and Mathews, Jang, Hami, and Stein (2004) did as well. If the putative disorder has four dimensions, which tend to share only 25%–31% of their variance with each other, then, by definition, individuals can be high on one dimension without being high on another dimension: Elevation in obsessive checking does not necessitate, for example, elevation in hoarding. The putative disorder is a combination of only moderately related constructs, and those constructs may have distinct genetic etiologies. OCD may not, on the basis of these findings, be a homogeneous psychological construct. To assign an individual a diagnosis of OCD may therefore be imprecise.

Posttraumatic stress disorder—Posttraumatic stress disorder (PTSD) is thought to be a distress-based disorder in hierarchical models (Watson, 2005). Its symptoms were shown to fall on four factors (Intrusions, Avoidance, Dysphoria, and Hyperarousal) by Simms, Watson, and Doebbeling (2002). Intercorrelations among the four ranged from .43 to .61, indicating substantial unshared variance in each factor. King, Leskin, King, and Weathers (1998) also found that a four-factor model (Reexperiencing, Effortful Avoidance, Emotional Numbing, and Hyperarousal) fit their 17-symptom clinical interview better than did any other model, including a single-factor model or a hierarchical model, in which an overall PTSD factor was thought to underlie the four factors. There is thus reason to question whether PTSD is best considered to be a theoretically coherent psychological entity. Clearly, identical PTSD symptom counts can refer to different symptom pictures. It may not be in patients' best interests to assign them a diagnosis that lacks clear meaning.

Schizotypal personality disorder—The apparent heterogeneity of some disorders according to the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychiatric Association, 1994) is not limited to Axis I disorders. Fossati et al. (2005) compared several different factor structures for the schizotypal personality disorder (SPD) criteria and found that a three-factor model (Cognitive-Perceptual, Interpersonal, and Disorganization) fit best. Intercorrelations among the three factors ranged from .14 to .63, again indicating substantial unshared variance in each factor. Here, too, individuals can be high on one factor but not on another, raising the question of the coherence of SPD as a psychological construct. Again in this case, the same quantitative symptom count could reflect different dysfunctional experiences.

It follows from these examples of heterogeneity that, for many disorders, the use of diagnostic status or a disorder score as either a predictor or a criterion in theory-testing studies will tend to produce unclear results. A score on depression, or a depression diagnosis, reflects scores on several constructs. To test a theory that Experience X is a risk factor for depression is to be imprecise. It may be the case that Experience X is a risk factor for one factor within depression but not for other factors. A proper test of that possibility requires assessment of the separate components of depression and examination of the association between Experience X and the target factor. If, instead, one used an overall depression score, one would risk missing the association altogether. If none of the other components of depression were related to Experience X, then the use of an overall depression score would essentially average the effect of the target factor with the irrelevant components, perhaps thus obscuring the importance of the target factor.

Perhaps more problematically, in such a situation one has to assume that the symptom count score reflects the same variable for each person, but that may well not be true. Individuals could have similar symptom counts yet different patterns of scores on individual constructs. When that is the case, the symptom count is not a coherent theoretical entity, and its correlation with measures of other constructs has unclear meaning. The same concerns apply when a disorder score is used as a predictor rather than as a criterion.

Of course, it may be true that some of these disorders represent true syndromes, in that some of the time the multiple dimensions stem from a common cause, even though the dimensions do not always share the same etiology (analogous to the flu example above). If that turned out to be true, then the validity of the syndrome description would nevertheless in part be a function of the precision with which each dimension within the syndrome had been defined and studied. However, the failure (to date) to find evidence for this possibility (Kupfer et al., 2002) suggests it is unlikely.

An additional advantage of a focus on homogeneous dimensions of dysfunction is this: Just as some diagnoses may be best understood as a set of moderately related constructs, it is also true that certain constructs are represented in the diagnostic criteria for many different disorders. Psychological constructs cut across disorders, and disorders combine separate psychological constructs. In light of this reality, there is an opportunity for improved clarity and better guidance to clinicians by describing dysfunction in terms of its homogeneous dimensions.

Having described this problem, we want to emphasize that we are not promoting the view that the *DSM* system is devoid of validity. On the contrary, the *DSM-IV* committee conducted and published 175 qualitative literature reviews and 48 additional empirical studies to provide the scientific underpinnings of their efforts (Widiger & Clark, 2000). In fact, each new version of the *DSM* has relied more heavily on the available science than did previous versions, and each new version has benefited from an ever-increasing body of scientific knowledge. The progress in psychiatric diagnosis is similar to progress in other diagnostic fields (Berg & Blackstone,

2006). Indeed, to a considerable degree, the rapid growth of psychopathology research has been facilitated by the development of a diagnostic nomenclature. Our argument is that the next step in this process may be to recognize the need to describe dysfunction along homogeneous dimensions and thus not be bound by a syndromal hypothesis that lacks validity. Doing so promises to further advance the validity of the diagnostic system, certainly in terms of its theoretical clarity and, we argue below, its parsimony and utility.

On the Comprehensiveness, Utility, and Parsimony of Using Homogeneous Constructs to Describe Psychopathology

We have presented an argument for describing psychopathology in terms of homogeneous dimensions of dysfunction, thus replacing the ill-fitting syndrome approach. We next consider three obvious concerns relevant to this proposal. First, can one describe psychopathology as comprehensively if one does not consider syndromes? That is, is an essential element of a given form of dysfunction likely to be missed if one does not consider full syndromes? Second, is description in terms of homogeneous dimensions useful? Does one lose the descriptive and communicative utility of the *DSM* approach? Third, is our approach parsimonious, or are we proposing an unwieldy, overly complex system that would require clinicians to assess across far too many dimensions of functioning? We address each issue in turn.

The comprehensive coverage of psychopathology with description in terms of homogeneous dimensions of dysfunction—Clinicians may feel that, in some cases, recognition of the existence of a syndrome is crucial to successful diagnosis. For example, psychopathy is considered, by some, to be a classic psychiatric syndrome. When clinicians observe antisocial behavior in a client, they are alert to the possibilities of a deceitful interpersonal style and limited capacity for empathy, among other attributes. If psychopathy were not a recognized syndrome, would clinicians be at greater risk of missing these additional characteristics of such clients?

To address this possibility, one must consider what an alternative client description, based on homogeneous dimensions of functioning, would consist of. Such a description was recently provided by Samuel and Widiger (2006). They compared diagnoses made with *DSM-IV* criteria with those made with the 30 personality trait scales from the NEO PI-R (the six facet scales for each of the five dimensions). They used three classic clinical cases, one of which was a 1.5-page case history of Ted Bundy (Bundy was not identified in the study). Using the *DSM-IV* criteria, 96% of clinicians diagnosed him with antisocial personality disorder, and 80% described the case as prototypic of that disorder. Dimensional diagnosis revealed that mean scores on the 30 traits within the FFM involved a description of him as lacking normal anxiety, self-consciousness, vulnerability, and warmth; as being nontrustworthy, not straightforward, not altruistic, not compliant, and not modest; and as unusually low in tender-mindedness. He was rated as unusually high in angry hostility, assertiveness, activity level, excitement seeking, competence, order, and achievement striving.

That personality description shows that clinicians did not appear to miss the essential elements of what has been thought of as psychopathic dysfunction. Indeed, for each of the three clinical cases in that study, the clinician participants rated the FFM description as superior to the *DSM-IV* diagnosis with respect to the FFM's capacity to describe all of the client's important personality difficulties (and on dimensions of utility, which we describe below; Samuel & Widiger, 2006). The success of this approach, at least in this study, is in part a function of the existence of comprehensive models of personality that have received extensive validation in the basic science literature (Costa & McCrae, 1995; Digman, 1990) and in the clinical literature (Clark, 2007; Widiger & Samuel, 2005; Widiger & Trull, 2007). When description is done in

terms of homogeneous dimensions of functioning, clinicians can take advantage of the extensive validity evidence in the literature.

Similar developments have occurred in the study of emotional disorders, as described in this special section by Brown and Barlow (2009). They present a compelling argument that valid description of anxiety and mood disorders would be enhanced through adoption of their model of homogeneous dimensions of affectivity.

The utility of disaggregation in psychopathology description—Samuel and Widiger (2006) reported evidence that clinicians found the FFM descriptions of psychopathology more useful, not just for describing all of a client's important personality problems, but also for communicating clearly with the client and others concerning the nature of his or her difficulties, for planning treatment, and for global personality description.

Another aspect of utility is the time taken to conduct a diagnostic assessment. Widiger and Lowe (2008) offered a practical proposal for personality disorder assessment in which a client is assessed across the 26 facets of personality thought to be most clinically relevant. They noted that assessment in this way takes approximately half the time necessary for a *DSM-IV* personality disorder assessment. Elsewhere in this special section, Mullins-Sweatt and Widiger (2009) provide additional proposals that speak to the efficiency of this form of description. It thus appears that in less time, one can obtain assessment data that clinicians find more useful.

In addition, it does not appear to be the case that a shift to our proposed form of description would require clinicians to learn new concepts or a new professional language. The FFM personality traits are readily accessible to clinicians, as are the affectivity dimensions involved in the Brown and Barlow (2009) proposal.

There are, of course, practical difficulties associated with a shift from the syndromal model to one based on homogeneous dimensions of functioning. First (2005) noted that such dramatic change would require retraining, cause administrative problems, complicate record keeping, disrupt some forms of research, and disrupt clinicians' ability both to apply past research to current clinical care and to communicate effectively with other mental health practitioners. First also recognized that none of these issues have to do with the validity of the science involved. As he noted, if the change provided improved validity and clinical applicability, it might be worth the disruption. At present, there is ample reason to believe that diagnosis along homogeneous dimensions would improve the validity and clinical applicability of psychiatric diagnosis.

The parsimony of disaggregation in psychopathology description—We believe that a focus on homogeneous dimensions of dysfunction would, in fact, provide improved parsimony over the current diagnostic system. At present, there are numerous syndromes in the *DSM*, and many of them share common dimensions of dysfunction. In the domain of personality disorders, description of psychopathology along homogeneous dimensions of personality is already quite well advanced. Researchers may be close to a consensus that domains of dysfunction in personality can be described in terms of four basic personality dimensions and their underlying facets (Widiger, Livesley, & Clark, 2009; Widiger & Simonsen, 2005; Widiger, Simonsen, Krueger, Livesley, & Verheul, 2005). That is, patterns of elevations across well-established personality traits can be used to describe the dysfunction currently described by the full set of personality disorders. There appears to be an advance in parsimony by targeting basic dimensions of personality dysfunction rather than trying to delineate multiple syndromes. And these suggestions of improved parsimony are not limited to personality disorders. There appears to have been important recent progress in identifying

dimensions of dysfunction shared by both personality disorders and Axis I clinical disorders (Krueger, 2002; Widiger & Simonsen, 2005).

To summarize this second section of the article, we maintain that when construct validation and theory testing are based on homogeneous constructs, the clarity of hypothesized and observed relationships among construct measures is advanced. We therefore recommend that psychopathology researchers study dysfunction with unidimensional or homogeneous constructs. This approach also provides increased comprehensiveness of client descriptions, increased utility, and perhaps increased parsimony of psychopathology description. In fact, the study of psychopathology along homogeneous dimensions of functioning is already quite advanced. In the field of personality disorders, descriptive models based on normal personality functioning exist and have been described by clinicians as more useful than the *DSM* system. Researchers have begun to show that a similar approach can be applied to Axis I disorders as well. Of course, nothing about this approach obviates the value of hierarchical descriptive models, and nothing about this approach precludes the presence of true syndromes of psychopathology.

Summary and Conclusion

We have argued for a significant shift in how clinical researchers approach the acquisition of scientific knowledge. Empirical data have been quite consistent with the possibility that terms that are routinely used in clinical inquiry, from *neuroticism* and *extraversion* to *depression* and *posttraumatic stress disorder*, do not in fact represent meaningful, cohesive psychological constructs; rather, they represent combinations of constructs. As a result, their use in investigations of the validity of constructs or theories is not recommended. To maximize the validity of findings from their inquiries, clinical scientists may be best served by conducting tests with measures of unidimensional, homogeneous constructs. Only measures of such constructs can represent coherent, causally active psychological entities.

Although this position represents a departure from common research practice, it represents a straightforward application of familiar psychometric principles. Clinical researchers are well positioned to focus on homogeneous measures of psychological entities. In a sense, we are only arguing that clinical researchers take their own psychometric theories seriously; they should avoid both complex items and complex scales, in order to best approximate human psychological functioning with their measures. Researchers have increasingly begun to apply this perspective in recent years. As a result, they have identified coherent, homogeneous dimensions of function and dysfunction, thereby positioning the field for significant advances in understanding psychopathology.

There are compelling reasons for clinical researchers to embrace this perspective. Should the authors of the *DSM-V* continue to presume the syndromal status of composite constructs, and thus define multidimensional entities as the proper object of study, the science of psychopathology research will develop more slowly than necessary. We believe one component of the next significant advance in the *DSM* process is to set aside the traditional but perhaps ill-fitting syndromal perspective and focus instead on homogeneous dimensions of dysfunction.

We believe that, at the most basic level, the approach we advocate will result in more sound tests of the validity of theories and the measures used to represent them. Progress in our field will be accelerated. We believe that the approach we are advocating, whether embraced by the *DSM-V* process or not, can contribute to more accurate descriptions of human psychological dysfunction and can therefore set the stage for more successful efforts to improve many individuals' quality of life.

Acknowledgments

Portions of this research were supported by National Institute on Alcohol Abuse and Alcoholism Awards R01 AA 016166 to Gregory T. Smith and R21 AA015218 to Denis M. McCarthy and National Institute on Drug Abuse Training Grant DA 007304 to Thomas Garrity for the training of Tamika C. B. Zapolski.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th. Washington, DC: Author; 1994.
- Bagby RM, Joffe RT, Parker JDA, Schuller DR. Re-examination of the evidence for the *DSM-III* personality disorder clusters. *Journal of Personality Disorders* 1993;7:320–328.
- Berg AT, Blackstone NW. Concepts in classification and their relevance to epilepsy. *Epilepsy Research* 2006;70(Suppl 1):11–19.
- Bollen K. Latent variables in psychology and the social sciences. *Annual Review of Psychology* 2002;53:605–634.
- Bollen K, Lennox R. Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin* 1991;110:302–314.
- Borsboom D, Mellenbergh GJ, van Heerden J. The theoretical status of latent variables. *Psychological Review* 2003;110:203–219. [PubMed: 12747522]
- Borsboom D, Mellenbergh GJ, van Heerden J. The concept of validity. *Psychological Review* 2004;111:1061–1071. [PubMed: 15482073]
- Brinkley CA, Newman JP, Widiger TA, Lynam DR. Two approaches to parsing the heterogeneity of psychopathy. *Clinical Psychology: Science and Practice* 2004;11:69–94.
- Brown TA, Barlow DH. A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: Implications for assessment and treatment. *Psychological Assessment* 2009;21 xxx–xxx.
- Bunnell CA, Winer EP. Lumping versus splitting: The splitters take this round. *Journal of Clinical Oncology* 2002;20:3576–3577. [PubMed: 12202656]
- Church T, Burke P. Exploratory and confirmatory tests of the Big Five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology* 1994;66:93–114. [PubMed: 8126653]
- Clark LA. Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology* 2007;58:227–257.
- Cleckley, H. *The mask of sanity*. St. Louis, MO: Mosby; 1941.
- Cooke DJ, Michie C. Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment* 2001;13:171–188. [PubMed: 11433793]
- Costa, PT.; McCrae, RR. *Revised NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources; 1992.
- Costa, PT.; McCrae, RR. The developing structure of temperament and personality from infancy to adulthood. In: Halverson, G.; Kohnstamm, G.; Martin, R., editors. *Stability and change in personality from adolescence through adulthood*. Hillsdale, NJ: Erlbaum; 1994. p. 139–150.
- Costa PT, McCrae RR. Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment* 1995;64:21–50. [PubMed: 16367732]
- Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychological Bulletin* 1955;52:281–302. [PubMed: 13245896]
- Cyders MA, Flory K, Rainer S, Smith GT. Prospective study of the integration of mood and impulsivity to predict increases in maladaptive action during the first year of college. *Addiction* 2009;104:193–202. [PubMed: 19149813]
- Cyders MA, Smith GT. Mood-based rash action and its components: Positive and negative urgency and their relations with other impulsivity-like constructs. *Personality and Individual Differences* 2007;43:839–850.
- Cyders MA, Smith GT. Clarifying the role of personality dispositions in risk for increased gambling behavior. *Personality and Individual Differences* 2008a;45:503–508. [PubMed: 19088857]

- Cyders MA, Smith GT. Emotion-based dispositions to rash action: Positive and negative urgency. *Psychological Bulletin* 2008b;134:807–828. [PubMed: 18954158]
- Cyders MA, Smith GT, Spillane NS, Fischer S, Annus AM, Peterson C. Integration of impulsivity and positive mood to predict risky behavior: Development and validation of a measure of positive urgency. *Psychological Assessment* 2007;19:107–118. [PubMed: 17371126]
- De Luca A, Botillo I, Sarkozy A, Carta C, Neri C, Bellacchio E. NFI gene mutations represent the major molecular event underlying neurofibromatosis-Noonan syndrome. *American Journal of Human Genetics* 2005;77:1092–1101. [PubMed: 16380919]
- Depue RA, Collins PF. Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences* 1999;22:491–569. [PubMed: 11301519]
- Digman JM. Personality structure: Emergence of the five factor model. *Annual Review of Psychology* 1990;41:417–440.
- Digman JM. Higher-order factors of the Big Five. *Journal of Personality and Social Psychology* 1997;73:1246–1256. [PubMed: 9418278]
- Edwards JR. Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods* 2001;4:144–192.
- Evenden J. Varieties of impulsivity. *Journal of Psychopharmacology* 1999;146:348–361.
- First MB. Clinical utility: A prerequisite for the adoption of a dimensional approach in *DSM*. *Journal of Abnormal Psychology* 2005;114:560–564. [PubMed: 16351379]
- Fischer S, Smith GT. Binge eating, problem drinking, and pathological gambling: Linking behavior to shared traits and social learning. *Personality and Individual Differences* 2008;44:789–800.
- Fischer S, Smith GT, Annus A, Hendricks M. The relationship of neuroticism and urgency to negative consequences of alcohol use in women with bulimic symptoms. *Personality and Individual Differences* 2007;43:1199–1209.
- Fischer S, Smith GT, Cyders MA. Another look at impulsivity: A meta-analytic review comparing specific dispositions to rash action in their relationship to bulimic symptoms. *Clinical Psychology Review* 2008;28:1413–1425. [PubMed: 18848741]
- Fischer, S.; Smith, GT.; Spillane, N.; Cyders, MA. Urgency: Individual differences in reaction to mood and implications for addictive behaviors. In: Clark, AV., editor. *The psychology of mood*. New York: Nova Science; 2005. p. 85-108.
- Fossati A, Citterio A, Grazioli F, Borroni S, Carretta I, Maffei C, Battaglia M. Taxonic structure of schizotypal personality disorder: A multiple-instrument, multi-sample study based on mixture models. *Psychiatry Research* 2005;137:71–85. [PubMed: 16226811]
- Frances, AJ.; First, MB.; Pincus, HA. *DSM-IV guidebook*. Washington, DC: American Psychiatric Press; 1995.
- Fukayama Y, Osawa M. To the epilepsy research 2006 supplement: Epileptic syndromes in infants and early childhood evidence-based taxonomy and its implication in the ILAE classifications. *Epilepsy Research* 2006;70:1–3.
- Georgi H, Glashow SL. Gauge theories without anomalies. *Physical Review D* 1972;6:429–431.
- Goldberg, LR. The structure of personality traits: Vertical and horizontal aspects. In: Funder, DC.; Parke, RD.; Tomlinson-Keasey, C.; Widaman, K., editors. *Studying lives through time: Personality and development*. Washington, DC: American Psychological Association; 1993. p. 169-188.
- Greven K. The problem of “lumping versus splitting”. *Gynecologic Oncology* 2005;99:527–529. [PubMed: 16269170]
- Hare, RD. *PCL-R technical manual*. Towanda, NY: Multi-Health Systems; 2003.
- Harpur TJ, Hakistan AR, Hare RD. Factor structure of the Psychopathy Checklist. *Journal of Consulting and Clinical Psychology* 1988;56:741–747. [PubMed: 3192791]
- Harpur TJ, Hare RD, Hakistan AR. Two-factor conceptualization of psychopathy: Construct validity and assessment implications. *Psychological Assessment* 1989;1:6–17.
- Hofstee W, de Raad B, Goldberg L. Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology* 1992;63:146–163. [PubMed: 1494982]

- Hogan, R.; Hogan, J. Hogan Personality Inventory manual. 2nd. Tulsa, OK: Hogan Assessment Systems; 1992.
- Hough, LM.; Schneider, RJ. Personality traits, taxonomies, and applications in organizations. In: Murphy, KR., editor. Individuals and behavior in organizations. San Francisco: Josey-Bass; 1995. p. 31-88.
- Jang K, Livesley W, Taylor S, Stein M, Moon E. Heritability of individual depressive symptoms. *Journal of Affective Disorders* 2004;80:125–133. [PubMed: 15207925]
- Jang KL, Livesley WJ, Vernon PA. Heritability of the Big Five personality dimensions and their facets: A twin study. *Journal of Personality* 1996;64:577–591. [PubMed: 8776880]
- Keller, L. Levels of selection in evolution. Princeton, NJ: Princeton University Press; 1999.
- King D, Leskin G, King L, Weathers F. Confirmatory factor analysis of the Clinician-Administered PTSD Scale: Evidence for the dimensionality of posttraumatic stress disorder. *Psychological Assessment* 1998;10:90–96.
- Kline, RB. Principles and practice of structural equation modeling. New York: Guilford Press; 2005.
- Kraepelin, E. Clinical psychiatry: A textbook for physicians (A Diefendorf, Trans.). New York: Macmillan; 1981. Original work published 1883
- Krueger, RF. Psychometric perspectives on comorbidity. In: Helzer, JE.; Hudziak, JJ., editors. Defining psychopathology in the 21st century: DSM–V and beyond. Washington, DC: American Psychiatric Association; 2002. p. 41-54.
- Krueger R, Markon K. Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology* 2006;2:111–133.
- Kupfer, DJ.; First, MB.; Regier, DE. Introduction. In: Kupfer, DJ.; First, MB.; Regier, DE., editors. A research agenda for DSM–V. Washington, DC: American Psychiatric Association; 2002. p. xv-xxiii.
- Leckman JF, Grice DE, Boardman J, Zhang H, Vitale A, Bondi C, et al. Symptoms of obsessive-compulsive disorder. *American Journal of Psychiatry* 1997;154:911–917. [PubMed: 9210740]
- Lynam DR, Widiger TA. Using a general model of personality to identify the basic elements of psychopathy. *Journal of Personality Disorders* 2007;21:160–178. [PubMed: 17492919]
- Markon K, Krueger F, Watson D. Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology* 2005;88:139–157. [PubMed: 15631580]
- Mathews CA, Jang KL, Hami S, Stein MB. The structure of obsessiveness among young adults. *Depression and Anxiety* 2004;20:77–85. [PubMed: 15468097]
- McCrae R, Zonderman A, Costa P, Bond M, Paunonen S. Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus procrustes rotation. *Journal of Personality and Social Psychology* 1996;70:552–566.
- McDermott, JJ., editor. The writings of William James. New York: Random House; 1967.
- McGrath RE. Conceptual complexity and construct validity. *Journal of Personality Assessment* 2005;112–124. [PubMed: 16171410]
- Miller J, Flory K, Lynam D, Leukefeld C. A test of the four factor model of impulsivity related traits. *Personality and Individual Differences* 2003;34:1403–1418.
- Morgan AJ. Editorial: ‘Splitting’ and ‘lumping’ reconciled? *Cell Biology International* 1997;21:617. [PubMed: 9778382]
- Mullins-Sweatt SN, Widiger TA. Clinical utility and DSM–V. *Psychological Assessment* 2009;21 xxx–xxx.
- Palmgreen, P.; Donohew, L. Handbook of drug abuse prevention: Theory, science, and practice. In: Sloboda, Z.; Bukoski, W., editors. Effective mass media strategies for drug abuse prevention campaigns. New York: Springer; 2003. p. 27-43.
- Paunonen SV. Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology* 1998;74:538–556.
- Paunonen SV, Ashton MC. Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology* 2001;81:524–539. [PubMed: 11554651]
- Persons J. The advantages of studying psychological phenomena rather than psychiatric diagnosis. *American Psychologist* 1986;41:1252–1260. [PubMed: 3813184]

- Petry N. Substance abuse, pathological gambling, and impulsiveness. *Drug and Alcohol Dependence* 2001;63:29–38. [PubMed: 11297829]
- Robins E, Guze SB. Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. *American Journal of Psychiatry* 1970;126:983–987. [PubMed: 5409569]
- Rogers, R. Diagnostic and structured interviewing. Odessa, FL: Psychological Assessment Resources; 1995.
- Samuel D, Widiger T. Clinicians' judgments of clinical utility: A comparison of the *DSM-IV* and five-factor models. *Journal of Abnormal Psychology* 2006;115:298–308. [PubMed: 16737394]
- Sarkozy A, Schirinzi A, Lepri F, Bottillo I, De Luca A, Pizzuti A, et al. Clinical lumping and molecular splitting of LEOPARD and NF1/NF1-Noonan syndromes. *American Journal of Medical Genetics Part A* 2007;143:1009–1011. [PubMed: 17366582]
- Saucier G. Replicable item-cluster subcomponents in the NEO Five-Factor Inventory. *Journal of Personality Assessment* 1998;70:263–276. [PubMed: 9697330]
- Schneider RJ, Hough LM, Dunnette MD. Broad-sided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior* 1996;17:639–655.
- Simms L, Watson D, Doebbeling B. Confirmatory factor analyses of posttraumatic stress symptoms in deployed and nondeployed veterans of the Gulf War. *Journal of Abnormal Psychology* 2002;111:637–647. [PubMed: 12428777]
- Smith GT. On construct validity: Issues of method and measurement. *Psychological Assessment* 2005;17:396–408. [PubMed: 16393005]
- Smith GT, Fischer S, Cyders MA, Annus AM, Spillane NS, McCarthy DM. On the validity and utility of discriminating among impulsivity-like traits. *Assessment* 2007;14(2):155–170. [PubMed: 17504888]
- Smith GT, Fischer S, Fister SM. Incremental validity principles in test construction. *Psychological Assessment* 2003;15:467–477. [PubMed: 14692843]
- Smith GT, McCarthy DM. Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment* 1995;7:300–308.
- Strauss ME, Smith GT. Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology* 2009;5:89–113.
- Turk DC. The potential of treatment matching for subgroups of patients with chronic pain: Lumping versus splitting. *Clinical Journal of Pain* 2005;21:44–55. [PubMed: 15599131]
- Watson D. Rethinking the mood and anxiety disorders: A quantitative hierarchical model for *DSM-V*. *Journal of Abnormal Psychology* 2005;114:522–536. [PubMed: 16351375]
- Watson D, Wu K. Development and validation of the Schedule of Compulsions, Obsessions, and Pathological Impulses (SCOPI). *Assessment* 2005;12:50–65. [PubMed: 15695743]
- Whiteside SP, Lynam DR. The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences* 2001;30:669–689.
- Whiteside SP, Lynam DR. Understanding the role of impulsivity and externalizing psychopathology in alcohol abuse: Applications of the UPPS Impulsive Behavior Scale. *Experimental and Clinical Psychopharmacology* 2003;11:210–217. [PubMed: 12940500]
- Whiteside SP, Lynam DR, Miller JD, Reynolds SK. Validation of the UPPS Impulsive Behavior Scale: A four-factor model of impulsivity. *European Journal of Personality* 2005;19:559–574.
- Widiger TA, Clark LA. Toward *DSM-V* and the classification of psychopathology. *Psychological Bulletin* 2000;126:946–963. [PubMed: 11107884]
- Widiger TA, Livesley JW, Clark LA. An integrative dimensional classification of personality disorder. *Psychological Assessment* 2009;21 xxx–xxx.
- Widiger TA, Lowe JR. A dimensional model of personality disorder: Proposal for *DSM-V*. *Psychiatric Clinics of North America* 2008;31:363–378. [PubMed: 18638640]
- Widiger T, Samuel D. Diagnostic categories or dimensions? A question for the *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition*. *Journal of Abnormal Psychology* 2005;114:494–504. [PubMed: 16351373]
- Widiger TA, Simonsen E. Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders* 2005;19:110–130. [PubMed: 15899712]

- Widiger TA, Simonsen E, Krueger R, Livesley J, Verheul R. Personality disorder research agenda for the *DSM-V*. *Journal of Personality Disorders* 2005;19:317–340.
- Widiger TA, Trull TJ. Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist* 2007;62:71–83. [PubMed: 17324033]
- Zuckerman, M. Behavioral expressions and biological bases of sensation seeking. Cambridge, England: Cambridge University Press; 1994.

Figure 1.

A depiction of one hierarchical organization of personality constructs. At the highest level are alpha and beta, at the next level are the five factors of the Big Five model of personality, and at the lowest level are the 30 specific trait or facet scales (six scales per factor; for simplicity's sake, we have not labeled each of the 30 facets). A = Agreeableness; C = Conscientiousness; N = Neuroticism; E = Extraversion; O = Openness to Experience.