Published in final edited form as:

Ind Organ Psychol. 2019 June ; 12(2): 157–162. doi:10.1017/iop.2019.31.

Constructs versus Measures in Personality and Other Domains: What Distinguishes Normal and Clinical?

Brenton M. Wiernik,

Department of Psychology, University of South Florida

Marina Bornovalova,

Department of Psychology, University of South Florida

Stephen E. Stark,

Department of Psychology, University of South Florida

Deniz S. Ones

Department of Psychology, University of Minnesota.

Psychopathology has long been recognized as dysfunction of normal psychological systems (Cloninger, 1987; Eysenck, 1947). Indeed, examination of psychological disorders is one of the avenues through which the structure of normal personality was discovered (Gibby & Zickar, 2008). Melson-Silimon and colleagues (2018) describe a collision course between the objective of accurate personnel assessment and the need for organizations to provide access for persons with mental health disabilities. Their alarm is misplaced. Frameworks for distinguishing normal and clinical assessments are already well-established, and the need to use different instruments or scoring methods for workplace versus clinical assessment is not unique to the personality domain. In this commentary, we highlight the critical distinction between *constructs* and their normal versus clinical *measurement* (cf. Dilchert et al., 2014) and demonstrate that normal and clinical personality measures have distinct psychometric properties, even while measuring the same underlying personality constructs. We also show that Melson-Silimon et al.'s concerns about similarity of normal and clinical constructs and their measurement applies to a wide variety of psychological constructs that are routinely assessed in organizational applications. We end by urging caution in interpreting normal and clinical personality measures and offer evidence-based guidance for personality assessment practice.

Personality

Clinical Constructs = Normal Constructs; Clinical Measures Normal Measures

Melson-Silimon et al. (2018) cite recent work from clinical psychology showing that personality disorders reflect maladaptive extreme low (for Extraversion, Conscientiousness, Agreeableness, or Emotional Stability) or high (Openness) levels of the Big Five personality traits (Markon et al., 2005). They suggest that this research exposes organizations using

normal measures of the Big Five personality traits or similar constructs to legal risks under the Americans with Disabilities Act (ADA), arguing that normal personality measures can be used to diagnose personality disorders. This suggestion reflects a confusion between the nature of psychopathology *constructs* and their assessment using psychopathology *measures*.

It may be the case that the personality psychopathology construct of “Disinhibition” reflects extremely low levels of the normal trait “Conscientiousness”, but this does not imply that Conscientiousness scales from the NEO PI-R, the Hogan Personality Inventory, the IPIP Big Five Markers, or other normal Conscientiousness measures can by themselves diagnose personality disorders. Normal and clinical personality measures are not interchangeable, and a scale designed for one purpose cannot be readily used for the other. Normal personality measures capture variations in generally healthy, adaptive responses to the environment. In contrast, well-designed clinical scales capture extreme trait levels that reflect disordered, maladaptive responses. To assess these extreme responses, clinical items are much more severe than normal items. For example, an item measuring normal depression (a facet of low Emotional Stability) is “Sometimes I feel completely worthless” (Costa & McCrae, 1992), whereas an exemplar item from a clinical depression scale is “During the past month, how much were you distressed by thoughts of ending your life?” (Derogatis & Melisaratos, 1983). Because clinical items are so extreme, they will often show substantial floor effects when administered to individuals without a psychological disorder. Conversely, because their content is insufficiently severe, normal personality items cannot distinguish people with disorders from those without.

From an item response theory (IRT) perspective, normal and clinical personality scales may assess the same latent traits, but they provide statistical information (reliable measurement) at much different trait levels. Normal personality scales tend to provide information in the range of ± 2 standard deviations from the general population mean. Well-designed clinical measures tend to provide information at much lower trait ranges, often being sensitive to individual differences -3 or -4 standard deviations below the mean or lower¹. Scales designed to tap one trait range can provide little reliable information about a different trait range.

This point is illustrated in Figure 1 using data on two Emotional Stability measures—one normal (the Stress Reaction scale of the Multidimensional Personality Questionnaire [MPQ]; Tellegen, 1990) and one clinical (the Depression items of the Brief Symptoms Inventory [BSI]; Derogatis & Melisaratos, 1983). The MPQ is a normal-range personality inventory widely used for research purposes and some workplace applications. The BSI is a short measure of psychological distress used in practice as a screening tool to identify potential problem areas for more focused diagnostic assessment. We estimated IRT item parameters for the two scales using responses from 1,046 MTurk workers (511 for the MPQ, 525 for the BSI) and 225 patients receiving treatment for substance use disorder.² We estimated IRT parameters separately for each scale, retaining one relatively non-severe item

¹These ranges are for clinical measures of Extraversion, Conscientiousness, Agreeableness, and Emotional Stability. For Openness, clinical psychosis measures are sensitive to individual differences in the Experiencing (sensory, experiential) aspect of Openness $+3$ to $+4$ standard deviations *above* the mean or higher (Grazioplene et al., 2016).

²IRT models were estimated using the *mirt* package (Chalmers, 2012, version 1.29) in *R* (R Core Team, 2018).

from the BSI³ in both item sets to use as an anchor to equate the scales for both measures. We then combined the item sets and estimated IRT information provided by each scale at different latent trait levels.

Figure 1 shows that the MPQ scale provides the most reliable assessment in a range from $\theta = -0.89$ to $+2.03$, whereas the clinical BSI scale provided reliable assessment from $\theta = -3.16$ to $+1.00$. Thus, the MPQ can discriminate in trait ranges reflecting high emotional stability to modestly high (but still normal-range) neuroticism. In contrast, the BSI scale can reliably discriminate healthy individuals from those with moderately severe clinical levels of depression. It is clear that the normal-range MPQ Stress Reaction scale provides no reliable information at clinically-relevant levels of depression. It is not a clinical scale, it cannot be used to diagnose psychological or mental disabilities, and it therefore is not a medical assessment.

Unity of Clinical and Normal Constructs is Not Unique to Personality Inventories

Normal personality traits and personality disorders are increasingly recognized as reflecting the same underlying trait dimensions (Dilchert et al., 2014; Markon et al., 2005). However, this is not unique to the personality domain. Most, if not all, constructs measured by personnel assessments capture the same underlying dimensions as clinical or medical instruments. Cognitive ability tests are widely used to predict employee training and job performance, but they are also used to detect a variety of clinical conditions, such as general learning disability, specific learning disabilities, ADHD, concussion, and cognitive decline (Kongs et al., 2000; McGrew et al., 2014; McLean et al., 2004; Mickes et al., 2007; Schatz et al., 2006). There is no question that these measures assess the same general and specific abilities. Normal and clinical cognitive assessments are distinguished not by the abilities they capture, but by the difficulty of the items and the norms used to interpret scores (McGrew et al., 2014). Similarly, employment interviews, assessment center exercises, and situational judgment tests are routinely used to assess, among other constructs, interpersonal skills (Barrick et al., 2009; Viswesvaran & Ones, 2018). Similar methods are also used to inform clinical decision making for psychological disorders such as autism spectrum disorder, schizophrenia, and borderline personality disorder (Edwards et al., 2002; First et al., 2016; Levine et al., 1997; Lord et al., 2000). Interviews, integrity tests, biodata measures, and credit and background checks frequently capture information about past counterproductive work behaviors, non-work deviant behaviors, and illegal drug use. Similar behaviors are also commonly used to diagnose antisocial personality disorder and other externalizing psychopathology (Krueger et al., 2002), all of which can be understood as extreme expressions of the same underlying disinhibited behavioral tendencies (Stanek et al., 2017). Even physical ability tests, such as measures of postural control, static strength, and speed, are used both in personnel selection (e.g., for military, protective services, and sports selection; Hoffman, 1999; Kuzmits & Adams, 2008) and for diagnosis of work disability, aging, and other medical conditions (Berg et al., 2009; El Fassi et al., 2013). Virtually any method for evaluating individual characteristics could be used for either clinical/medical

³“During the past month, how much were you distressed by: Feeling blue?”

assessment to diagnose disorders or to compare individual differences in normal functioning for personnel applications.

The key factor that distinguishes normal and medical assessments is not the similarity of the underlying constructs, but rather the extremity of the individual responses that they can detect and the norms used to evaluate them. If, following the reasoning offered by Melson-Silimon et al. (2018), most normal range personality measures cannot be used for personnel assessment because they assess the same constructs as clinical diagnostic personality scales, this would preclude virtually all individual assessment of any construct on the same grounds.

Implications of Construct Unity for Personality Assessment Practice

Melson-Silimon et al. (2018) conclude with a set of drastic recommendations that would serve to discard virtually all existing I-O personality assessment research and practice. The most alarmist recommendation is that organizations should “Avoid personality tests that assess constructs closely related to PDs and PID-5 dimensions, including “off-the-shelf” tests developed with the ADA in mind, tests known to assess PDs, “dark side” traits, and normal personality traits that are highly correlated with PDs.” (p. 34). This list encompasses all personality measures used in workplace assessment. The Big Five traits occupy the center of the personality trait hierarchy (John et al., 2008; Stanek & Ones, 2018). All personality measures, even those developed from other theoretical perspectives (Pace & Brannick, 2010), will show substantial correlations with Big Five measures and, because they are defined as clinical extremes of the Big Five, likely with measures of the DSM-5 personality disorder dimensions as well.

But superficial content similarity to or even substantial correlations with clinical instruments are not the standards which make a measure of personality (or any other construct) a medical diagnostic tool based on EEOC (2000) guidance. Normal personality scales typically used in employee selection and screening do not have enough sensitivity at extreme trait levels to indicate a medical diagnosis. That is, they are not designed to *nor can they reliably* diagnose or reveal mental disorders. If a personality scale provides no statistical information at clinically-relevant trait ranges and does not have norms that indicate how scores can be used to diagnose mental disorders, then it is not a medical examination.

Organizations wishing to measure personality variables as part of their pre-employment selection and assessment systems can prevent identification of psychopathology and psychological disorders by exercising caution and heeding the following evidence-based advice:

1. Ensure personality scales used are supported by a job analysis (Raymark et al., 1997), as well as validity evidence.
2. Do not use measures designed to assess psychological disorders or mental health (e.g., the MMPI, the BSI, the PID-5) in pre-offer personnel assessments.
3. Do not use extreme items drawn from clinical instruments designed to assess psychological disorders as part of personnel assessment scales.

4. Interpret scores using occupational norms developed for the specific assessment purpose (e.g., selection, development), not clinical norms or diagnostic cut scores.

Developing bespoke personality instruments, avoiding the Big Five constructs, or disregarding personality traits altogether is neither necessary nor likely to be a productive path for reliable, valid, and unbiased assessment of employee motivational characteristics.

Conclusion

Normal and clinical personality scales are distinguished in terms of *where* on the trait distribution they provide reliable information, and they cannot be substituted for one another. This distinction parallels long understood distinctions between ability measures, interviews, and other assessments that can also be designed to capture normal, exceptional, or dysfunctional functioning depending on the design purpose and test properties. Normal personality scales by themselves can provide little useful information for the diagnosis of clinical disorders and cannot be considered medical instruments.

References

- Barrick MR, Shaffer JA, & DeGrassi SW (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology*, 94(6), 1394–1411. <https://doi.org/10/cwc7mj> [PubMed: 19916651]
- Berg K, Wood-Dauphine S, Williams JI, & Gayton D (2009). Measuring balance in the elderly: Preliminary development of an instrument. *Physiotherapy Canada*. <https://doi.org/10/dc8nr8>
- Chalmers RP (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29. <https://doi.org/10/gckfrg>
- Cloninger CR (1987). A systematic method for clinical description and classification of personality variants: A proposal. *Archives of General Psychiatry*, 44(6), 573–588. <https://doi.org/10/frm495> [PubMed: 3579504]
- Costa PT, & McCrae RR (1992). *Professional manual for the NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory (NEO-FFI)* [Technical manual]. Odessa, FL: Psychological Assessment Resources.
- Derogatis LR, & Melisaratos N (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine*, 13(3), 595–605. [PubMed: 6622612]
- Dilchert S, Ones DS, & Krueger RF (2014). Maladaptive personality constructs, measures, and work behaviors. *Industrial and Organizational Psychology*, 7(1), 98–110. <https://doi.org/10/gckfs5>
- Edwards J, Jackson HJ, & Pattison PE (2002). Emotion recognition via facial expression and affective prosody in schizophrenia: A methodological review. *Clinical Psychology Review*, 22(6), 789–832. <https://doi.org/10/fj6g2v> [PubMed: 12214327]
- EEOC. (2000). Enforcement guidance on disability-related inquiries and medical examinations of employees under the Americans with Disabilities Act (ADA) (Notice No. 915.002). Retrieved from Equal Employment Opportunity Commission website: <https://www.eeoc.gov/policy/docs/guidance-inquiries.html>
- El Fassi M, Bocquet V, Majery N, Lair ML, Couffignal S, & Mairiaux P (2013). Work ability assessment in a worker population: Comparison and determinants of Work Ability Index and Work Ability score. *BMC Public Health*, 13(1), 305 <https://doi.org/10/gbc2jj> [PubMed: 23565883]
- Eysenck HJ (1947). *Dimensions of personality*. London, United Kingdom: Routledge & Kegan Paul.
- First MB, Williams JBW, Karg RS, & Spitzer RL (2016). *Structured Clinical Interview for DSM-5 Disorders: Clinician Version (SCID-5-CV)*. Arlington, VA: American Psychiatric Association.

- Gibby RE, & Zickar MJ (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of Psychology*, 11(3), 164–184. <https://doi.org/10/dzk7ss> [PubMed: 19048975]
- Grazioplene RG, Chavez RS, Rustichini A, & DeYoung CG (2016). White matter correlates of psychosis-linked traits support continuity between personality and psychopathology. *Journal of Abnormal Psychology*, 125(8), 1135–1145. <https://doi.org/10/f9bvhc> [PubMed: 27819473]
- Hoffman CC (1999). Generalizing physical ability test validity: A case study using test transportability validity generalization and construct-related validation evidence. *Personnel Psychology*, 52, 1019–1041.
- John OP, Naumann LP, & Soto CJ (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues In John OP, Robins RW, & Pervin LA (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). New York: Guilford Press.
- Kongs SK, Thompson LL, Iverson GL, & Heaton RK (2000). *Wisconsin Card Sorting Test-64 card version (WCST-64)* [Technical manual]. Lutz, FL: Psychological Assessment Resources.
- Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WG, & McGue M (2002). Etiologic connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology*, 111(3), 411–424. <https://doi.org/10/czm9gg> [PubMed: 12150417]
- Kuzmits FE, & Adams AJ (2008). The NFL Combine: Does it predict performance in the National Football League? *Journal of Strength & Conditioning Research*, 22(6), 1721 <https://doi.org/10/c9p88h> [PubMed: 18841077]
- Levine D, Marziali E, & Hood J (1997). Emotion processing in borderline personality disorders. *Journal of Nervous and Mental Disease*, 185(4), 240 <https://doi.org/10/fnshk3> [PubMed: 9114809]
- Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, ... Rutter M (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223. <https://doi.org/10/c3gpcq> [PubMed: 11055457]
- Markon KE, Krueger RF, & Watson D (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, 88(1), 139–157. <https://doi.org/10/dprbt5>
- McGrew KS, LaForte EM, & Schrank FA (2014). *Woodcock-Johnson IV* [Technical Manual]. Retrieved from Riverside website: <https://www.wjsscore.com/Files/WJIVTechnicalManual.PDF>
- McLean A, Dowson J, Toone B, Young S, Bazanis E, Robbins TW, & Sahakian BJ (2004). Characteristic neurocognitive profile associated with adult attention-deficit/hyperactivity disorder. *Psychological Medicine*, 34(4), 681–692. <https://doi.org/10/cqrc33> [PubMed: 15099422]
- Melson-Silimon A, Harris AM, Shoenfelt EL, Miller JD, & Carter NT (2018). Personality testing and the Americans with Disabilities Act: Cause for concern as normal and abnormal personality models are integrated. *Industrial and Organizational Psychology*. Retrieved from: <https://www.siop.org/journal/12T/focal2a.pdf>
- Mickes L, Wixted JT, Fennema-Notestine C, Galasko D, Bondi MW, Thal LJ, & Salmon DP (2007). Progressive impairment on neuropsychological tasks in a longitudinal study of preclinical Alzheimer's disease. *Neuropsychology*, 21(6), 696–705. <https://doi.org/10/dzjsw9> [PubMed: 17983283]
- Pace VL, & Brannick MT (2010). How similar are personality scales of the “same” construct? A meta-analytic investigation. *Personality and Individual Differences*, 49(7), 669–676. <https://doi.org/10/ccjmkmm>
- R Core Team. (2018). R: A language and environment for statistical computing. Retrieved from: <http://www.r-project.org/>
- Raymark PH, Schmit MJ, & Guion RM (1997). Identifying potentially useful personality constructs for employee selection. *Personnel Psychology*, 50(3), 723–736. <https://doi.org/10/bzvvy5t>
- Schatz P, Pardini JE, Lovell MR, Collins MW, & Podell K (2006). Sensitivity and specificity of the IMPACT Test Battery for concussion in athletes. *Archives of Clinical Neuropsychology*, 21(1), 91–99. <https://doi.org/10/ctz8f2> [PubMed: 16143492]

- Stanek KC, & Ones DS (2018). Taxonomies and compendia of cognitive ability and personality measures relevant to industrial, work, and organizational psychology In Ones DS, Anderson N, Viswesvaran C, & Sinangil HK (Eds.), *The SAGE handbook of industrial, work and organizational psychology* (2nd ed., Vol. 1, pp. 366–407). 10.4135/9781473914940.n14
- Stanek KC, Ones DS, & McGue M (2017). Counterproductive behaviors: Relations across life domains, etiology, and implications for applied practice. *International Journal of Selection and Assessment*, 25(2), 111–124. <https://doi.org/10/gckf9r>
- Tellegen A (1990). *Manual for the Multidimensional Personality Questionnaire*. Minneapolis, MN: University of Minnesota Press.
- Viswesvaran C, & Ones DS (2018). Non-test methods and techniques used in employee selection In Ones DS, Anderson N, Viswesvaran C, & Sinangil HK (Eds.), *The SAGE handbook of industrial, work and organizational psychology* (2nd ed., Vol. 1, pp. 451–473). 10.4135/9781473914940.n16

Figure 1.

IRT information plot for *normal* (MPQ Stress Reaction, solid blue line) and *clinical* (BSI Depression, dashed red line) Emotional Stability scales. Normal scale, MPQ Stress Reaction, provides maximum information at $\theta = +0.53$ and information > 2.0 from -0.89 to $+2.03$. Clinical scale, BSI Depression, provides maximum information at $\theta = -1.31$ and information > 2.0 from -3.16 to $+1.00$.