Industrial and Organizational Psychology (2019), **12**, 177–183
doi:10.1017/iop.2019.35

CAMBRIDGE UNIVERSITY PRESS

COMMENTARY

Tilting at windmills and improving personality assessment practices

Neil D. Christiansen[1]*, Peter A. Fisher[2], Chet Robie[2], and Stuart Quirk[3]

[1]Central Michigan University, [2]Wilfrid Laurier University, and [3]Magnolia Counseling Associates
*Corresponding author. Email: chris1nd@cmich.edu

Melson-Silimon, Harris, Shoenfelt, Miller, and Carter (2019) bring attention to the use of normative personality inventories for pre-employment testing in light of the Americans with Disabilities Act (ADA) and recent developments in research on personality. Although they review some of the legal guidelines and case precedent, they concede that no court case has yet taken issue with personality tests that did *not* contain items based on the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1942) and offer some suggestions for employers who wish to conduct personality assessments without potential conflict with the ADA. In this commentary, we discuss why concerns raised in the focal article may not prove to be problematic after all and offer advice for employers based on best practice in the area.

Do integrated models suggest existing normative personality tests are medical exams?

Although the focal article discusses advances in integrative models of personality that include both dimensions of normative personality (five-factor model; FFM) and clinical personality (DSM-5), it is important to realize that the ADA focuses on *measures* and *assessment results*, rather than *theoretical models*. Camara and Merenda (2000) summarized the Equal Employment Opportunities Commission (EEOC) guidelines and case precedent with regard to whether a test would be considered a medical screen and concluded that the decision hinges on answers to four questions:

1. Is the test *designed* to reveal impairment or physical or mental health (such diagnosis is the primary intended use for personality tests developed with clinical normative populations)? (emphasis in original)
2. Is the test normally given in a medical setting (for example, a health care professional's office)?
3. Is the test administered or interpreted by a health care professional or someone trained by a health care professional?
4. Is medical equipment used?

Personality tests designed specifically for pre-employment screening purposes would likely yield a "no" answer to all of these questions (Schmit & Ryan, 2013), as would any normative trait inventory not developed or used for clinical purposes. In contrast, the MMPI was specifically developed to assist psychologists and other mental health professionals with differential diagnosis and the development of treatment plans (Hathaway & McKinley, 1942). Changes to theoretical models that psychologists use in research are unlikely to alter the answers to the four questions for

© Society for Industrial and Organizational Psychology 2019.

https://doi.org/10.1017/iop.2019.35 Published online by Cambridge University Press

existing personality inventories that predate the integrative models or for those that are developed thereafter. However, using common sense with regard to how the results of the personality assessments are maintained in the organization and ensuring that there is written policy precluding medical professionals from accessing them should render these concerns moot.

It may still be useful to consider the scenario where the plaintiff argues that a normative test is effectively a proxy for an assessment of psychopathology, such as the MMPI or Personality Inventory for DSM-5 (PID-5). Although there is evidence of some commonality between measures of normative personality and clinical personality, this does not suggest the measures of each are redundant or isomorphic (or more properly anti-isomorphic). For example, Quirk, Christiansen, Wagner, and McNulty (2003) regressed each of 28 MMPI-2 scales (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) onto all 30 of the facets of the NEO PI-R (Costa & McCrae, 1992) and found that the normative personality trait scales explained an average of approximately 40% of the variance in the MMPI scale scores (average Multiple $R = .63$ for all 30 facets).

To directly estimate the overlap between corresponding measures of each model, we conducted a bare-bones meta-analysis of the FFM/PID-5 convergent validity estimates based on eight effect sizes located in the literature (Fowler et al., 2017; Gore, 2013; Helle, Trull, Widiger, & Mullins-Sweatt, 2017; Kajonius, 2017; Pocnet, Antonietti, Handschin, Massoudi, & Rossier, 2018; Sleep, Hyatt, Lamkin, Maples-Keller, & Miller, 2018; Watson, Stasik, Ro, & Clark, 2013).[1] As can be seen in Table 1, the sample-weighted mean convergent validity estimate between Openness to Experience and Psychoticism ($r = .18$) was the smallest, whereas that for Emotional Stability and Negative Affectivity ($r = –.61$) was the largest in absolute terms.[2] Omitting the mean correlation between Openness and Psychoticism (that does not directly correspond to the relationship to any DSM personality disorder), the average uncorrected correlation is less than .50 in the expected negative direction. This suggests considerable overlap but is not nearly large enough to assume one inventory could substitute for the other after scores were inverted.

Beyond this convergence, Quirk et al. (2003) found that there was evidence that normative personality patterns were related to actual PD symptomatology. This is consistent with the relationships cited in the focal article from the meta-analysis of Saulsman and Page (2004), who also found substantial correlations between the FFM and the PDs from the DSM-IV; however, again, none of the mean correlations exceeded .50. Importantly, when explaining symptoms related to Axis II Personality Disorders (PDs), Quirk et al. (2003) found that each type of personality assessment explained incremental variance in the PD symptoms beyond the other (increasing the Multiple *R*s an average of more than .20). Use of scores from each personality test also increased the correct diagnostic classifications for all of the PDs beyond accurate classification for only one type of test. If the two inventories were predicting the same symptomatology, no such incremental validity or improved classification should be noted. Taken together, we do not share the concern of Melson-Silimon et al. (2019) that the correlations between the FFM and DSM-5 measures are strong enough that the courts might classify normative personality tests as clinical measures, and they are certainly not interchangeable in predicting PD symptomatology. However, it does suggest that the results of each type of assessment might be used in tandem to make better diagnostic decisions in clinical settings.

Even if normative personality inventories are helpful for making PD diagnoses because there are distinct patterns of extreme trait elevations, regular use in employment settings is unlikely to conflict with the ADA. Consider the following life activities that might be asked about by a hiring

[1]See Schmidt and Hunter (2015) for approaches to bare-bones meta-analysis. One study (Helle, Trull, Widiger, & Mullins-Sweatt, 2017) contributed two effect sizes because they used two different operationalizations of the FFM. All were published studies except for a master's thesis (Gore, 2013).

[2]Sample size did not account for the majority of the variance in any of the validity estimates, suggesting that moderators are likely present.

https://doi.org/10.1017/iop.2019.35 Published online by Cambridge University Press

manager during an employment interview or appear on a normative personality inventory in the assessment of orderliness, sociability, and trust:

- Balancing a checking account
- Keeping living space neat and organized
- Maintaining lasting relationships
- Receiving praise and approval from family

These life activities might well be impaired by a variety of clinical disorders, including PDs. Questions regarding these activities could therefore also be asked during a clinical interview or on a psychopathology scale to gauge functioning on disinhibition, antagonism, and detachment. Do emerging integrative models that organize the latter traits into the same framework as the former set suggest that using questions about these activities prior to a conditional offer of employment might be construed by the courts as providing a medical examination?

Consider again the four questions that must be answered in the affirmative for a psychological test to potentially be considered a medical examination according to Camara and Merenda (2000). We believe existing normative personality inventories that inquire about the above activities would still yield a "no" to all of those questions. The intention during the development of the normative inventory was not to gauge impairment or mental illness but rather to assess normative, functional behavior. We therefore believe that just because integrated models include traits assessed by the NEO PI-R, and given that there might be increased use of these inventories by clinicians, regular use in employment settings is not likely to conflict with the ADA. If normative personality test results were being used more *inside* medical settings than *outside*, that might change.

This line of reasoning also applies to the research discussed that extreme scores on normative traits might be associated with adaptive behavior in a curvilinear fashion, such as whether very high levels of conscientiousness reflect obsessive-compulsive tendencies (Carter, Guan, Maples, Williamson, & Miller, 2016). If the normative personality test were not developed to assess impairment or mental illness, and no medical professionals were involved in the testing or interpretation of the results, use in an employment screening is consistent with the guidelines set forth by the EEOC with regard to medical examinations and the ADA. Consider a biodata assessment intended to assess the trait of core self-evaluations in job applicants. If it were found in other research that the biodata assessment identified a disproportionate number of individuals with narcissistic PD, it still was not developed with the intention of identifying individuals with a disability, nor were the results used for this purpose. As long as employers take proper care with regard to access of the results by medical professionals that might use the results to identify individuals with clinical disorders, we do not see liability risk to the employer under ADA.

The area of the most obvious concern from the perspective of the ADA is that individuals scoring very low on the normative emotional stability scale would also be likely to score high on many of the clinical scales and be more likely to meet diagnostic criteria across multiple disorders. However, the relationships between clinical scales, PD symptoms, and normative emotional stability are ubiquitous enough that low emotional stability scores by themselves would be unlikely to be useful for *differential diagnosis* (Quirk et al., 2003). When traits related to this FFM dimension are assessed in applicants, it would therefore be critical to demonstrate that any candidates who were screened out had not received an offer because of the job relevance of the scores and not any impairment related to a disability. Toward this end, we offer the following suggestions toward best practice.

Improving personality assessment practices in organizations

Standing by our principles

The focal article by Melson-Silimon et al. (2019) advocates that employers using personality tests for hiring always conduct a job analysis and collect data for a local criterion-related validity study.

https://doi.org/10.1017/iop.2019.35 Published online by Cambridge University Press

We agree, especially because this is recommended practice based on the *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, 2018). However, personality tests, in particular, are often used in human resource functions such as hiring and promotion in ways that appear to radically depart from these practices. For example, Christiansen, Tett, and Robie (2011) surveyed HR managers who report using personality tests for hiring decisions and found that approximately 75% of the organizations used *all of the trait scales* included on multidimensional tests in decision making based on a "profile comparison" method.

Ideally, the job analysis and criterion-related validity study would winnow the number of traits from an inventory that get used for decision making, given that on average only half of the scales on a given personality inventory are likely to demonstrate any useful level of validity (Simpson, Christiansen, & Dawson, 2017). In addition, many decisions seem to be made primarily based on narrative summary reports, and actual cutoff scores are only utilized by approximately 5% of the organizations using these inventories for hiring (Christiansen et al., 2011). Whether a noncompensatory model with multiple cutoff scores is utilized or trait scale scores are combined and a single cutoff score is used, we believe that best practice is to use cutoff scores after a job analysis and criterion-related validity study have winnowed the number of scale scores to be included in selection decision making. Use of cutoff scores, especially when they were derived based on reference to a meaningful performance threshold, should assist in demonstrating job relevance of decisions and therefore reduce the chance that a personality test is challenged as a medical examination.

Not all job analyses are equal

It is also important that the results of the job analysis provide strong evidence that the normative traits in the personality inventory are related to success on the critical work activities for the job. This would be helpful with any dispute that the assessment is being used for discriminatory purposes. O'Neill, Goffin, and Rothstein (2013) identified three basic approaches to personality-oriented job analysis (POJA):

1. A traditional hybrid method where a task list is developed along with selection requirements in terms of knowledge, skills, and abilities (KSAs) important for success identified based on those activities. A testing plan then links the KSAs to the traits in the personality inventory.
2. A trait-based method where job experts review the task list and then rate whether the traits from the inventory would have an overall negative or positive effect on performance, along with how strongly related the traits should be to success.
3. A behaviorally based method where general work activities have been linked to traits *a priori* by applied psychologists with expertise in the area. Job experts then rate how important each activity is to success in the job with higher ratings on relevant behavior indicating the trait is necessary for performance.

We note that none of these methods necessarily includes the additional step of linking the traits used in the personality inventory to specific tasks on the job. Similar to a task–KSA linkage matrix that is sometimes used in job analysis, this involves presenting job experts with the specific tasks of the job and asking them to identify if the traits from the inventory are important for success. This makes the job relatedness obvious if the trait scales being used have been shown to be related to the most critical work activities and may therefore forestall legal challenges. Even if trait scores have been empirically shown to be related to job performance in a validity study, including them in decision making may be questionable if they cannot be logically argued to hinder or help important work activities.

What's in a name?

Finally, we agree that making the case that a personality inventory is being used for a legitimate employment practice, rather than as a medical examination, may depend on how the traits

https://doi.org/10.1017/iop.2019.35 Published online by Cambridge University Press

Table 1. Meta-analysis of convergent validity estimates from measures of the five-factor model and personality inventory for DSM-5

Convergent validity	Mean *r*	SD_r	% Var.	90% CI	80% CrI
E/DE	−.50	.07	36.2	[−.54, −.46]	[−.56, −.44]
A/AN	−.43	.20	5.8	[−.55, −.32]	[−.66, −.20]
C/DI	−.41	.27	2.8	[−.57, −.25]	[−.73, −.09]
ES/NA	−.61	.08	18.1	[−.57, −.66]	[−.53, −.70]
O/P	.18	.08	47.5	[.14, .23]	[.12, .24]

Note. $N = 2{,}755$, $k = 8$. E = Extraversion; DE = Detachment; A = Agreeableness; AN = Antagonism; C = Conscientiousness; DI = Disinhibition; ES = Emotional Stability; NA = Negative Affectivity; O = Openness to Experience; P = Psychoticism. Mean *r* = sample size weighted mean observed correlation; SD_r = standard deviation of the observed correlation; % Var. = percentage of variance accounted for by sample size; 90% CI = 90% confidence interval; 80% CrI = 80% credibility interval.

are labeled and defined. Because of this, naming personality trait scales in maladaptive terms such as volatility and withdrawal may be inviting problems that might be avoided if stability and self-regard could serve instead. Of course, this may also involve defining the traits as how frequently someone engages in behavior at the positive end of the trait's behavioral domain more so than the negative. As such, using trait names such as anxiety or depression that appear in diagnostic criteria for clinical disorders is almost asking for trouble, and we agree with the focal article that it might be best to altogether avoid personality measures that explicitly use PDs for dimensional definition or "dark-side" traits that may be seen as subclinical (Furnham, Richards, & Paulhus, 2013; Hogan & Hogan, 2001).

The focal article carries this one step further and discusses developing inventories that only inquire about activities on the job and that involve only behavior that occurs at work rather than general tendencies or activities that might occur while not at work. We fundamentally agree. It is known that contextualizing personality inventories has benefits with regard to applicant reactions and validity (e.g., Hunthausen, Truxillo, Bauer, & Hammer, 2003). However, research has too often focused on shifting frame of reference by adding work "tags" to items from a normative personality inventory. We believe more inventories need to be developed that make the job context obvious by changing a general question ("I have a good imagination") to one that could only happen on the job ("I prefer jobs that require innovation"). Research confirms that such "fully contextualized" personality measures are more predictive and face valid than less contextualized measures (cf. Robie, Risavy, Holtrop, & Born, 2017).

In conclusion, we do not believe that it matters if extreme scores on normative measures are related to PDs as long as they were not designed with the purpose of identifying disabilities and simple precautions are taken with regard to medical professionals never being part of the process. To be proactive and avoid complaints of any kind with the EEOC, organizations using personality inventories are advised to consider four processes that will lead to best practice:

1. Complete both a job analysis (ideally with the trait scales used for applicant assessment being linked to specific tasks from the job) and a criterion-related validity study to narrow the number of trait scales used and ensure job relatedness.
2. Set cutoff scores on the inventory that are criterion referenced, based on an established performance threshold to be sure that decisions are systematically made only on the job-related trait scales.
3. Avoid labeling personality trait scales in maladaptive terms, especially anything resembling PDs found in the DSM.
4. Use context-specific personality inventories that only have items related to behaviors found in the workplace.

https://doi.org/10.1017/iop.2019.35 Published online by Cambridge University Press

Author ORCIDs. Peter A. Fisher, 0000-0002-6845-109X

References

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory—2: Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.

Camara, W. J., & Merenda, P. F. (2000). Using personality tests in preemployment screening: Issues raised in *Soroka v. Dayton Hudson Corporation*. *Psychology, Public Policy, and Law*, **6**(4), 1164–1186. doi: 10.1037/1076-8971.6.4.1164

Carter, N. T., Guan, L., Maples, J. L., Williamson, R. L., & Miller, J. D. (2016). The downsides of extreme conscientiousness for psychological well-being: The role of obsessive compulsive tendencies. *Journal of Personality*, **84**, 510–522. doi: 10.1111/jopy.12177

Christiansen, N. D., Tett, R., & Robie, C. (2011, May). *Problems and pitfalls in organizational uses of personality inventories*. Paper presented at the European Congress of Work and Organizational Psychology, Maastricht, Netherlands.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI–R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Fowler, J. C., Patriquin, M. A., Madan, A., Allen, J. G., Frueh, B. C., & Oldham, J. M. (2017). Incremental validity of the PID-5 in relation to the five factor model and traditional polythetic personality criteria of the DSM-5: DSM-5 personality trait domains. *International Journal of Methods in Psychiatric Research*, **26**(2), e1526. doi: 10.1002/mpr.1526

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The dark triad of personality: A 10 year review. *Social and Personality Psychology Compass*, **7**, 199–216. doi: 10.1111/spc3.12018

Gore, W. L. (2013). The DSM-5 dimensional trait model and the five factor model. *Theses and Dissertations—Psychology*, **12**. https://uknowledge.uky.edu/psychology_etds/12

Hathaway, S. R., & McKinley J. C. (1942). *Manual for the Minnesota Multiphasic Personality Inventory*. Minneapolis, MN: University of Minnesota Press.

Helle, A. C., Trull, T. J., Widiger, T. A., & Mullins-Sweatt, S. N. (2017). Utilizing interview and self-report assessment of the five-factor model to examine convergence with the alternative model for personality disorders. *Personality Disorders: Theory, Research, and Treatment*, **8**(3), 247–254. doi: 10.1037/per0000174

Hogan, R., & Hogan, J. (2001). Assessing leadership: A view from the dark side. *International Journal of Selection and Assessment*, **9**, 40–51. doi: 10.1111/1468-2389.00162

Hunthausen, J. M., Truxillo, D. M., Bauer, T. N., & Hammer, L. B. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology*, **88**, 545–551. doi:10.1037/0021-9010.88.3.545

Kajonius, P. J. (2017). The Short Personality Inventory for DSM-5 and its conjoined structure with the common five-factor model. *International Journal of Testing*, **17**(4), 372–384. doi: 10.1080/15305058.2017.1309421

Melson-Silimon, A., Harris, A. M., Shoenfelt, E. L., Miller, J. D., & Carter, N. T. (2019). Personality testing and the Americans with Disabilities Act: Cause for concern as normal and abnormal personality models are integrated. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, **12**(2), 119–132.

O'Neill, T. A., Goffin, R. D., & Rothstein, M. G. (2013). Personality and the need for personality-oriented work analysis. In N. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 226–253). New York, NY: Routledge.

Quirk, S. W., Christiansen, N. D., Wagner, S. H., & McNulty, J. (2003). On the usefulness of measures of normal personality for clinical assessment: Evidence of the incremental validity of the Revised NEO Personality Inventory. *Psychological Assessment*, **15**(3), 311–325. doi: 10.1037/1040-3590.15.3.311

Pocnet, C., Antonietti, J.-P., Handschin, P., Massoudi, K., & Rossier, J. (2018). The many faces of personality: The DSM-5 dimensional and categorical models and the five-factor model. *Personality and Individual Differences*, **121**, 11–18. doi: 10.1016/j.paid.2017.09.005

Robie, C., Risavy, S. D., Holtrop, D., & Born, M. P. (2017). Fully contextualized, frequency-based personality measurement: A replication and extension. *Journal of Research in Personality*, **70**, 56–65. doi: 10.1016/j.jrp.2017.05.005

Saulsman, L. M. & Page, A. C. (2004). The Five-Factor Model and personality disorder empirical literature. *Clinical Psychology Review*, **23**(8), 1055–1085. doi: 10.1016/j.cpr.2002.09.001

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Los Angeles, CA: Sage.

Schmit, M. J., & Ryan, A. M. (2013). Legal issues in personality testing. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 525–542). New York, NY: Routledge.

Simpson, D. A., Christiansen, N. D., Dawson, K. (2017, April). Occupational differences in the validity of narrow personality traits. Paper presented in J. Johnson (Ed.), *Don't take quotes or personality assessment validities out of context*. Symposium presented at the 32nd Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

https://doi.org/10.1017/iop.2019.35 Published online by Cambridge University Press

Sleep, C. E., Hyatt, C. S., Lamkin, J., Maples-Keller, J. L., & Miller, J. D. (2018). Examining the relations among the DSM-5 alternative model of personality, the five-factor model, and externalizing and internalizing behavior. *Personality Disorders: Theory, Research, and Treatment*, **9**(4), 379–384. doi: 10.1037/per0000240

Society for Industrial and Organizational Psychology, Inc. (2018). *Principles for the validation and use of personnel selection n procedures* (5th ed.). Arlington Heights, IL: Author. https://www.apa.org/ed/accreditation/about/policies/personnel-selection-procedures.pdf

Watson, D., Stasik, S. M., Ro, E., & Clark, L. A. (2013). Integrating normal and pathological personality: Relating the DSM-5 trait-dimensional model to general traits of personality. *Assessment*, **20**(3), 312–326. doi: 10.1177/1073191113485810

Cite this article: Christiansen N.D., Fisher P.A., Robie C., and Quirk S. (2019). Tilting at windmills and improving personality assessment practices. *Industrial and Organizational Psychology* **12**, 177–183. https://doi.org/10.1017/iop.2019.35

https://doi.org/10.1017/iop.2019.35 Published online by Cambridge University Press