Evaluating the Accuracy of Personality Tests: A Reflective Report

Table of Contents

Introduction to Personality Testing

Personality testing has become an important area of modern psychology and is applied extensively in clinical diagnosis, employee selection, and personal development. The primary basis of these tests is self-report questionnaires that are supposed to capture an individual's enduring characteristics. Although the popularity of the Big Five derives from their empirical support and ability to be administered quickly, there are many criticisms of the reliability and validity of personality assessments based on the Big Five.

On the one hand, the advantage of personality testing is that it can evaluate broad traits or traits that the human being is accustomed to, considering that, to a large extent, they define human behaviour. Items of Big Five personality test such as tests based on robust factor analytic techniques and have shown the acceptable levels of internal consistency and short term test–retest reliability have been used (Goldberg, 1993; John, et al, 2008). These tools are based on solid theories of personality that maintain that essential characteristics such as extraversion or conscientiousness do not vary much over time. These tests are attractive for application and research due to their stability.

Nevertheless, measuring personality is a complex process. A significant problem with this is that self-reporting methods depend partly on the individual's self-awareness and honesty. Socially desirable responding (the tendency to give socially preferable answers in response to survey questions) is a frequent source of warning of bias on self-report measures (as pointed out by Mischel 1968). This is especially problematic when we realize our responses may have evaluative consequences, so we underreport less socially desirable traits (Crowne & Marlowe, 1960). Further complicating interpretation of results are response styles like acquiescence (responding agreeably to statements irrespective of their content).

One of the other critical issues is the construct validity of the personality test. Despite this, many instruments supposedly measuring 'real' parts of personality (Funder, 1991) are accused of oversimplifying very complex human activities into brief and static categories. As an example, although the Big Five model divides the traits into five major domains, the minority of the variations in the person's personality are not covered by this model (Soto & John, 2017). Furthermore, many tests are based on theoretical assumptions that are not universally accepted;

for instance, debates rage about whether personality traits are stable or somewhat susceptible to large amounts of fluctuations throughout time (Roberts & Mroczek, 2008).

However, personality assessment is even more difficult in a cultural context. Many popular tests have been developed and normed on Western, Educated, Industrialized, prosperous, and democratic (WEIRD) populations, leading many to question their generalizability to others of disparate backgrounds (Henrich et al., 2010). When such tests are applied across more than one cultural setting, they may not consider cultural nuance, and as a result, a particular personality trait may be misinterpreted or misclassified.

In the end, personality testing promises to give insight into individual differences, but limitations to this are necessary. Costs of bias, oversimplification, and the cultural specificity of standardized measures set the limit to their benefits. Thus, researchers and practitioners face the challenge of critically evaluating these instruments in light of both empirical evidence and personal experience so the conclusions made from such tests will … be both [reliable and] valid.

Brief Introduction to the Specific Test

I have used the Big Five personality test, the OCEAN model, to evaluate it for these purposes. A five-dimensional test measures five broad dimensions of a personality: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Goldberg, 1993). The Big Five instrument is often used as a self-report questionnaire containing about 40 to 100 items. The participant rates their agreement using a Likert-type scale with a series of statements.

The Big Five is one of personality psychology's best empirically supported personality frameworks (John, Naumann & Soto, 2008, developed through lexical and factor analytic studies). It is a parsimonious and comprehensive approach to understanding personality, leading to its popularity in academic research and practical applications. However, the Big Five test can't escape criticism, much less its widespread acceptance. Others dispute that the broad trait categories of the model can account for the nuance of personality or the effect of situational factors (Soto & John, 2017). However, its good psychometric properties and use in many cross-cultural studies still give it weight and practicality.

Evaluation of the Test Based on My Experiences as a Participant

The Big Five personality test was new to me, so I wanted to try it but simultaneously wondered whether there was any genuine value in it. The structured questionnaire format encouraged me to base my initial impressions on a plain and easy questionnaire. Yet, as I worked through these items, certain things about the test rang familiar and started to generate questions about its general reliability.

My overriding concern was that transient mood states would affect my responses. For example, when answering items on extraversion, I was indeed checked out on my current social mood and recent interaction. I remember thinking, 'I probably would have answered differently if I was more energetic or inclined socially.' Research showing that self-reported similarities among people can vary significantly with mood (Roberts & Mroczek, 2008) supports this observation. Thus, the Big Five dimensions may appear stable because of the testing environment rather than as an index of enduring personality traits.

Another thing that impressed me was the habit of giving socially desirable answers. Surprisingly, I also typically subconsciously gravitated towards responses that painted me in positives – in terms of agreeableness and conscientiousness, for example. This is consistent with Crowne and Marlowe's (1960) finding that social desirability bias is a significant shortcoming of self-report measures. While promised to be anonymous, there seems to be a built-in pressure to be present as myself the best way I can, all of which raises questions of the validity of the results. The problem is that such biases undermine the reliability of the test and jeopardize its validity as a means of personality testing.

I also have noticed that when the Likert scales had a forced choice aspect, the ability to express nuances of my personality was sometimes compromised. Specifically, while I tend to regard myself as moderately open to new experiences, my openness was also measured to some degree that the available response options may not capture. This limitation is not unique to my knowledge; several studies have pointed out that the binary or limited nature of some response options in many personality tests causes the loss of some subtle information (Podsakoff et al., 2003). A more flexible or adaptive format might also provide a better measure of the individual difference and thus increase overall assessment accuracy.

Further reflecting, I began to doubt whether the theoretical foundations of the Big Five model were paved in conditions fit for the complexity of human personality. Although the model relies on broad trait categories, it is a helpful framework, but some variation within each trait domain may be missed. One example is the trait neuroticism, which encompasses the full range of emotional experience from mild anxiety to severe mood instability in the test. However, the test aggregates this into a single, summary score. Thus, aggregation can mask significant differences that could be important from both a research and applied perspective. I share Funder's (1991) critique of reductionism in personality assessment; such words as a purer approach may provide more valuable conclusions.

I could identify the strengths of the Big Five test despite its limitations. It's not all that common and empirically supported well enough to earn some credibility in practice that few other personality measures can make the same claim. However, I still have concerns about the applicability of the test to others, given that the test has been shown to hold across multiple cultural contexts (Henrich, Heine, & Norenzayan, 2010). In addition, as the model is simple, it is easy for other people to access, including myself, and it allows me to get a basic knowledge of our personality traits without requiring much training in psychometrics.

One electronic positive aspect of the Big Five test is that it may be a tool for self-reflection. The questionnaire allowed me to introspect and think of some of my behaviours that I might have overlooked. While not a substitute for an excellent in-depth psychological assessment, this reflective process can offer some value to aid personal growth and increased self-awareness. While comparing my responses to the descriptive literature on personality, I found both congruences and incongruences in my self-perception and the theoretical constructs of the Big Five. Reflective practice in psychology (Schön, 1983) is a dialogue between personal experience and academic theory.

Reflections in the light of the foregoing render me believe that the Big Five personality responsiveness is suitable as a framework for understanding individual variation, yet several reasons limit increasing accuracy. There are numerous reasons why what should be considered an accurate measure of one's personality might not necessarily reflect what is reported in a personality questionnaire: transient mood states, social desirability biases, and other inherent constraints of the response format. Additionally, within the broad Big Five categories, some subtleties may be

necessary for a complete understanding of human behavior. This evidence, informed by my personal experiences and the supporting literature, underscores the need for more and more refinement in the field of personality assessment.

Some of these limitations, however, have been proposed to be alleviated by additional methods, such as informant reports or behavioral observations, integrated (Vazire, 2010). Pairing self-report data with external perspectives may enable researchers to gain a more complete picture of personality, such as the experience of personality and the way the individual appears to others. As a multimethod approach, this may prove to be the way of the future of personality assessment – using each method's strengths to bolster each mode's shortcomings.

Conclusion

As I think back on the Big Five personality test, I understand that the test was a good and supported tool of personality psychology but isn't perfect. All of this tracks the concerns of countless researchers concerned with my personal experience: influenced by transient mood effects, social desirability biases, and the constraints of forced choice responses. However, even with these problems, the Big Five model carries a high degree of influence due to its empirical standing and as an initial point of self-exploration. Adaptive testing formats and the merging of many assessment methods for the better assessment of the richness of human personality may be the avenue for future improvements in personality assessment. Overall, this refleActive exercise has made me better aware of the limits and potential of personality testing and has made me a more discerning bystander to its results.

References

Cheung, F. M., van de Vijver, F. J. R., & Leong, F. T. L. (2011). Toward a new approach to the study of personality in culture. *American Psychologist, 66*(7), 593–603.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349–354.

Funder, D. C. (1991). Global traits: A few keys to personality. *Psychological Inquiry, 2*(1), 59–83.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26–34.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61–83.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). Guilford Press.

Mischel, W. (1968). *Personality and assessment*. Wiley.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical literature review and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.

Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science, 17*(1), 31–35.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*(2), 281–300.