# Assessment in the Digital Age: Some Challenges for Test Developers and Users

**Prof. Stephen A. Woods**

## Key Digested Message

*How should psychometrics specialists (test developers and users) respond to the challenges of the digital age? A broad challenge that faces the discipline of psychometrics is to avoid being rooted in old ways of thinking about testing, whilst simultaneously ensuring that key principles of best practice in the science of assessment are maintained and applied to the new methodologies of testing. This article explores key issues in this focal challenge.*

## Introduction and Context

The main motivation for this article has been my experiences of the recent past years in which I have encountered new technologies in assessment. These include symposia at EAWOP and DOP conferences where new approaches such as game-based assessments have attracted strong opinion and debate about how such techniques should be evaluated. Then, in my practice with Aston Business Assessments (ABA), and as an academic at Surrey Business School, I have been frequently contacted by start-up companies in the assessment space, who are thinking innovatively, and in some examples, radically about the way in which data can be used to draw conclusions about psychological attributes. They have challenged my own thinking about the nature of psychological measurement in the digital age. Finally, it is my view that we have a strong ethical obligation to ensure we stay in the wider professional conversation about psychological assessment as the digital economy grows. Recent debates about data privacy, and revelations about the potential role of profiling in shaping public opinion for example, only serve to underline why psychologists need to adapt and to guide practice, attending to issues of assessment effectiveness and ethics.

This article is an opinion paper and will explore three substantive issues to illustrate challenges of assessment in the digital age:

1. The need to conceptualise assessments flexibly, as modular kits rather than fixed instruments.
2. Consideration of how reliability and validity methodology is applied in situations where assessment data are not conventionally collected.
3. Ensuring the relevance of testing for the new kinds of jobs that the digital economy enables.

## 1. Designing Assessment for Flexible Application

Psychometric testing is built on historical foundations of classical ‘instruments’; tests for which the development lead time was typically long, requiring extensive trialling and data collection. Production of materials for administration and scoring in hardcopy indicated the expected longevity of published instruments, a set up that prompted criticism of the field’s capacity to innovate.

Digital innovation now means that development time has been substantially reduced. For example, platforms such as MTurk and Prolific provide access to trial data in days, enabling fast design of questionnaires and surveys. Online presentation of materials also means there is no reason for psychometrics to be conceptualised as 'ready-assembled'. Rather, individual components of tests can be compiled in modules for construction into various forms to meet the needs of different contexts. In the area of personality assessment, for example, items and scales can be reconfigured based on normative data to create short parcels or composites directed to performance demands (such as competency scoring in personality inventories; e.g. see Woods & Anderson's Periodic Table of Personality; 2016).

Treating assessment components flexibly in this way is not new. The International Personality Item Pool (Goldberg, 1999) was designed exactly on this premise. Most test developers utilise item banking to help construct different tests based on client needs. The Trait Personality Inventory (ABA, 2011), for example, has been conceptualised as a flexible taxonomy, which has a standard form, but which has also been configured in multiple ways for different clients and assessment needs. Each configuration requires analyses of the psychometric properties of the resultant scales and components, yet this is possible with large databanks.

Designing assessments as modular kits will enable psychometric developers to be more responsive to methods of digital service creation. Digital start-ups grow organically, testing products with users, incorporating elements that users want and need. Insistence on the part of psychologists that our assessment technologies are fixed and non-configurable will prevent our participation in the development of new HR analytic technology.

**2. Psychometric Evaluation: Reliability and Validity**

The focus of psychometrics on conventional testing methods has arguably led to an overly simplified mental model of reliability and validity in which 'rules of thumb' or consensus benchmarks for quality are applied without critical consideration. A good example is the concept of construct validity. My discussions with entrepreneurs and digital developers have often highlighted new ways in which personality trait or psychological data are captured. In recent months, concepts of AI-based analyses of video interviews and written content (e.g. postings, emails) are just two examples. Assuming that participants had consented to this use of their data, what would convince us as psychologists that these methods have effectively measured, for example, a construct like Neuroticism?

Convention would dictate that convergence with an existing measure at around 0.70 or greater would be necessary. Yet, critical consideration of this mental model quickly highlights shortcomings. For instance, such convergence benchmarks assume methodological equivalence (i.e. converging scores on one self-report survey with another). Multi-method comparisons such as self-other agreement in survey ratings are not held against similar expectations. Rather the self- and other-ratings are treated as indicators of the construct (one self-perceptual, the other observation-based), between which convergence is expected to a reasonable degree. There is no assumption made that the two should be *equivalent* to demonstrate construct validity. We should likewise not reject outright the validity of new assessment methodologies simply because they do not conform to our conventional expectations of how construct validity evidence is evaluated. Rather, construct validity needs

to be critically appraised as evidenced by an accumulated set of indicators (Woods & West, 2014).

An additional risk of focusing on construct validity, is that matters of criterion validity are potentially neglected. Classical meta-analyses of the validities of selection methods are rooted in data collected decades before the introduction of digital methods. There are few published validation studies, and no meta-analyses of the validities of assessments delivered through digital platforms.

**3. Jobs in the Digital Economy**

In a forthcoming paper, colleagues and I re-examined theories of the impact of work on identity processes, focusing on unconventional employment, including the gig economy (see Selenko, Berkers, Carter, Woods, Otto, Urbach & De Witte, 2018). Our discussions that led to the paper were motivated by the need to be inclusive in the way that the field develops.

The rise of the gig economy demonstrates how digital platforms also impact the environments in which psychometrics are applied. For instance, how would we position the role of psychometrics as making an impact on people and businesses that operate in gig-economy? Who is the primary user of psychometric data, the worker (in order to self-assess into gig-economy careers) or the company that engages them (as an indicator of the potential of the worker)? I have yet to encounter a published peer-reviewed study on the criterion effects of individual differences on performance in these kinds of work environments, and so the evidence base from which to work in answering these questions is limited.

Potential routes to move into new work contexts are illustrated however, by looking at another emergent area of digital entrepreneurship, internet start-ups. In these micro-businesses, success factors include the composition of the team of founders. In this context, it is highly probable that personality composition also makes an impact, given what we know about the benefits of diversity when tasks are complex (Guillaume, Dawson, Otaye-Ebede, Woods & West, 2017). Psychometrics could have a key potential role in determining the potential success and failure of digital micro-business start-ups, if methods and tools are developed and applied in ways that enable them to be accessed and understood.

**Summary and Conclusion**

The three highlighted areas in this article are designed to illustrate and reflect how the digital economy might impact on the practice of psychometrics and assessment. These reflections represent one way of considering the issues, others may likely take different and alternative views. However, what is clear is that psychometric test developers, test users and psychologists must engage in critical reflection and debate to ensure that our approaches remain innovative and relevant in the digital age. In doing so, we will ensure that our scientific influence on assessment practice is maintained and embedded in new assessment technologies, and in the emergent environments that the digital economy creates.

## References

Aston Business Assessments Ltd (2011). *Trait: Manual and User Guide.* Birmingham UK; https://www.astonassessments.com

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, *7*(1), 7-28.

Guillaume, Y. R., Dawson, J. F., Otaye-Ebede, L., Woods, S. A., & West, M. A. (2017). Harnessing demographic differences in organizations: What moderates the effects of workplace diversity?. *Journal of Organizational Behavior*, *2*(38), 276-303.

Selenko, E., Berkers, H., Carter, A., Woods, S.A., Otto, K., Urbach, T., & De Witte, H. (2018). On the dynamics of work identity in atypical employment: Setting out a research agenda. *European Journal of Work and Organizational Psychology.*

Woods, S. A., & Anderson, N. R. (2016). Toward a periodic table of personality: Mapping personality scales between the five-factor model and the circumplex model. *The Journal of Applied Psychology*, *101*(4), 582-604.

Woods, S. A., & West, M. A. (2014). *The psychology of work and organizations 2e*. Cengage Learning EMEA.

*This opinion piece was originally written for and published in the BPS Assessment and Development Matters Journal in May 2018.*