BIROn - Birkbeck Institutional Research Online

Universal Design in Psychometric Testing

Universal Design (UD) is the concept that infrastructure and services should be conceived with abilities and diversity in mind, in order to reduce barriers for individuals. It is a key tenet of the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD, 2006), and has led to the introduction of dropped kerbs and step-free access to buildings across much of the globe. Universal Design has also been applied to 'Assistive Technology' (AT) (García-Betances et al., 2016; Van Laarhoven et al., 2018) and software design for everyday use. The rise of speech-to-text and text-to-speech software being added to common word processing and video conferencing software is a UD feature, as is growth in alternative-text (alt-text) options on social media and packages such as Microsoft PowerPoint and PDF format files. Below, we define each of the seven UD principles in turn, interpreting the descriptions provided by the Center for Universal Design (Rickerson, 2009; The Center for Universal Design, 1997). These descriptions are based on a Chapter written for the Society of Industrial and Organizational Psychology in their handbook on Neurodiversity at work (Doyle, 2021). We then discuss the implication of UD for psychometric test use and design, summarised in Table One, before summarising the next steps for practice and research.

Principle one: "Equitable Use"

The aim of Equitable Use is to avoid segregating disabled people and requiring them to ask for adjustments, which is stigmatising and results in fewer people coming forward for help. Notably those less likely to ask for help are often those additionally marginalised by other protected characteristics such as gender and race.

Principle two: "Flexibility in Use"

The Flexibility in Use principle requires an acknowledgement that not everyone will approach a task in the same way, and that infrastructure needs to accommodate difference. For example, increasing the speed of delivery on a video or podcast.

Principle three: "Simple and Intuitive Use"

Simple and Intuitive Use requires us to avoid unnecessary complexity and repetition, ensuring that information is laid out clearly and intuitively. With the design of many processes and tools, the designers become the experts and many easily forget how someone might misinterpret intention or be confused by the details without an overview, headings, visual signposting and more. Language accessibility can be improved by use of the Flesch-Kincaid scale, a grading system which uses the average word length, sentence length and paragraph length to estimate the reading level required to understand text (Flesch, 1948; Kincaid et al., 1981). The Flesch-Kincaid scale is built into the advanced grammar checks of widely available programs such as Microsoft Word.

Principle four: "Perceptible Information"

Perceptible Information refers to sensory modalities and acknowledges that some people might find reading, listening, and moving difficult as a result of disability. This principle encourages us to ensure more than one sensory modality for dissemination and access of information.

Principle five: "Tolerance for Error"

Tolerance for Error acknowledges that some people are more likely to make mistakes on the first or second time of operation / engagement, they may wish to go back and correct, they may not get the process, even when they can understand the content. By building tolerance for error into infrastructure, these people are not inadvertently penalised.

Principle six: "Low Physical Effort"

The principle of Low Physical Effort is targeted at predicting and minimizing physical effort, on account of the fact that physical strength is normally distributed and also affected by disability. The lower the effort required, the more people will have access.

Principle seven: "Size and Space for Approach and Use."

Size and Space for Approach and Use focuses on the adaptation of environments and workspace. Neurodiversity research, for example, has shown a need to consider noise, temperature, protection of personal space and privacy, visual stimulus and movement, smells and more (Shulamite et al., 2015; Whitby, 2018).

Table One: Universal Design Applied to Psychometric Test Design

Principle	Test Design	Test Instructions	Test Administration
Equitable Use	Compatibility with assistive technology	Compatibility with assistive technology	Ability for the test to be used with wheelchair users, people with sensory impairments and cognitive disabilities
Flexibility in Use	Speed and order of testing. The ability to take breaks.	Instructions can be given at multiple points, in a variety of formats.	Speed and order of testing. The ability to take breaks.
Simple and Intuitive Use	User experience factored into the design	Flesch-Kincaid score appropriate to the testee	Ability to explain and report results using a language level appropriate to the testee
Perceptible Information	Compatibility with assistive technology	Can be accessed verbally or written, braille or sign language options	Adequate debriefing, making sure reports are accessible and there is no time limit on questions
Tolerance for Error	Return to previous questions to correct errors	The ability to request repeats of instructions	For testers to notice where mistakes relate to understanding the process rather than the content
Low Physical Effort	Compatibility with assistive technology	All instructions and test stimuli for an item audible / visible at the	Plenty of breaks, splitting the test over two sessions

		same time to avoid the need for flicking between screens	
Size and Space for Approach and Use	Test can be delivered in a variety of settings, including seated and standing	Instructions are	Noting any disability that may affect concentration or access to testing centres

What does this mean in practice?

There are clear rules provided by test publishers regarding the conditions, order and instructions that should be provided during testing. Some of these are necessary for the measurement of the construct, for example we cannot repeat working memory test items as we are directly measuring the ability to retain the instruction. However, many tests have not adequately accommodated disabilities and, as a community, we need to advocate for disabled testees by exercising our consumer power. Ask before you buy about compatibility with assistive technology, for example. Further, if we do not make these adjustments, we may be measuring the wrong construct. Should situational judgment or verbal comprehension tests be dependent on reading or working memory? If the test administration involves reading complex scenarios or multiple-choice grammar, we may be measuring literacy or processing speed when the target is long-term memory. This is relevant for many professional exams, including those administered by the Medical Royal Colleges and the College of Policing. This is where the provision of extra time can be justified, even though the research on effectiveness is mixed (Holmes & Silvestri, 2019). Testing professionals should exercise judgment and ensure that any influences of visible and non-visible disablement are clearly stated to avoid incorrect interpretation on behalf of the client / testee. Here are some examples of accommodations I have made whilst using the Weschler Adult Intelligence Scale (WAIS: Weschler, 2008):

- Starting from the first item, rather than the fourth, in order to ensure that the testee understands the way a question should be answered and to reduce anxiety for people affected by cognitive disabilities.

- Completing the testing over two sessions, to avoid fatigue compromising results.

- Answering questions about what a test is for and what it might relate to in context to provide a big picture overview for those that need the landscape before they can engage in the detail.

- Pro-rating 2 sub-tests to form an index score even though a third subtest was attempted, to mitigate the impact of confusion, inaccessible instructions and fatigue.

- I have also printed out reading materials from literacy assessments (Wechsler, 2017) in larger font with double spacing to make them more accessible for dyslexics and people with visual impairments.

These accommodations require skill on behalf of the test administrator. You need to have good observation skills and also be able to cross reference information you know from the testee's background and history. For example, I once assessed a Dyspraxic woman who had been refused adjustments in education because her report indicated that all her scores were below the 20th percentile. This was interpreted as her having developmental disabilities and being unsuitable for higher education rather than a specific learning difference or neurodivergence. However, her GCSEs included an A grade and five grades above C. Her BTEC was a distinction. Her ability to converse with me in the background interview and independently follow directions to the appointment did not indicate a global development or learning disability. Knowing that often Dyspraxic people struggle to follow instructions, I made the accommodation of going through the verbal comprehension lower-level items in the WAIS in order for her to 'get' how to approach the question with very simple material. Following this, she achieved a standard score of 13 for verbal comprehension, slightly above

average, which was much more commensurate with her life history. When we reflected, she explained that she had been confused by the instructions, overwhelmed by the process and found the educator austere and hostile. Admittedly some of these issues could have been resolved by improving standards of communication, but testing professionals need to appreciate that for some testees the impact of rapport and anxiety is more extreme than for others.

Similarly, I have observed that many testees with Attention Deficit and Hyperactivity Disorder (ADHD) miss easy items early in the subtest because they are not concentrating, and then ace the last five. Now, anyone scoring full marks at the top end of a WAIS subtest should have a high IQ, but when the points are added up, they might only obtain a raw score of 30 out of 60, which would be more mediocre. This needs careful framing and unpicking during the debrief.

What does this mean for research?

We are sorely lacking research on the impact of disabilities on psychometric testing. In my past twenty years as a testing professional I have observed these patterns many times, and indeed my neurodiversity literate colleagues concur, but we still lack empirical evidence which would justify flexibility in interpretation and delivery as standard practice. Further, as the use of Artificial Intelligence in the application of testing increases, human interaction is increasingly out of the loop and caution is advised on any decision-making protocol where human administrations cannot influence outcomes by being in the loop or 'on the loop' (Bankins, 2021). Problems become compounded for disabled people who are rarely considered when designing video interviewing (consider those with facial disfigurement, tics, stroke patients) or online test use (consider Dyslexic, Blind and Deaf people) (Nugent et al., 2020). Situational judgement tests have been found to be discriminatory for Autistic people

by Employment Appeal Tribunal, meaning that this is now settled UK law[1]. My own research suggests a complex picture of interaction between testing administration and neurodivergence, more than failing to understand the process. Some neurodivergent people are highly skilled at spotting ambiguities and small errors that even the test designers failed to notice, which stops them in their tracks and slows them down (Doyle & Waseem, 2022), yet could be an important quality in a workplace. What if it is insight and attention to detail that requires extra time, rather than neurodivergent people needing help because they find testing difficult? Both options are possible, but we have previously assumed that extra time was deficit led rather than associated with strengths. This is an urgent matter for our profession and we need to develop a research agenda to explore the ethical use of testing when so many are potentially disadvantaged by the format and delivery. There's a PhD in there for anyone interested!

References

Bankins, S. (2021). The ethical use of artificial intelligence in human resource management: A decision-making framework. *Ethics and Information Technology*, *23*(4), 841–854. https://doi.org/10.1007/s10676-021-09619-6

Doyle, N. (2021). Adapting Other Internal Resources to a Neurodiverse Workforce. In S. Bruyere & A. Collela (Eds.), *Society of Industrial and Organizational Psychology: New Frontiers Series, Neurodiversity*. American Psychological Association.

Doyle, N., & Waseem, U. (2022). Using Clean Language interviewing to explore the lived experience of neurodifferent job applicants. In H. Cairns-Lee, J. Lawley, & P. Tosey (Eds.), *Clean Language Interviewing*. Emerald Publishing Ltd.

---

[1] https://www.gov.uk/employment-appeal-tribunal-decisions/the-government-legal-service-v-ms-t-brookes-ukeat-0302-16-rn

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221–233. https://doi.org/10.1037/h0057532

García-Betances, R. I., Cabrera-Umpiérrez, M. F., Ottaviano, M., Pastorino, M., & Arredondo, M. T. (2016). Parametric cognitive modeling of information and computer technology usage by people with aging-and disability-derived functional impairments. *Sensors (Switzerland)*, *16*(2). https://doi.org/10.3390/s16020266

Holmes, A., & Silvestri, R. (2019). Extra Time or Unused Time? What Data from a College Testing Center Tells Us About 50% Extra Time as an Accommodation for Students with Learning Disabilities. *Psychological Injury and Law*, *12*(1), 7–16. https://doi.org/10.1007/s12207-019-09339-9

Kincaid, J. P., Aagard, J. A., O'Hara, J. W., & Cottrell, L. K. (1981). Computer readability editing system. *IEEE Transactions on Professional Communication*, *24*(1), 38–42. https://doi.org/10.1109/TPC.1981.6447821

Nugent, S. E., Jackson, P., Scott-Parker, S., Partridge, J., Raper, R., Bakalis, C., Shepherd, A., Mitra, A., Long, J., Maynard, K., & Crook, N. (2020). *Recruitment AI has a Disability Problem: Questions employers should be asking to ensure fairness in recruitment* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/emwn5

Rickerson, N. (2009). Universal Design. In I. Soderback (Ed.), *International Handbook of Occupational Therapy*. Springer.

Shulamite, A. G., Hernandez, L., Tottenham, N., Krasileva, K., Bookheimer, S. Y., & Daprettor, M. (2015). Neurobiology of Sensory Overresponsivity in Youth With Autism Spectrum Disorders. *Journal of the American Medical Association Psychiatry*, *72*(8), 778–786. https://doi.org/10.1001/jamapsychiatry.2015.0737

The Center for Universal Design. (1997). *The Principles of Universal Design, Version 2.0.*

United Nations Convention on the Rights of Persons with Disabilities, 1 (2006).

Van Laarhoven, T., Carreon, A., Bonneau, W., & Lagerhausen, A. (2018). Comparing Mobile

Technologies for Teaching Vocational Skills to Individuals with Autism Spectrum Disorders

and/or Intellectual Disabilities Using Universally-Designed Prompting Systems. *Journal of*

Autism and Developmental Disorders, *48*(7), 2516–2529. https://doi.org/10.1007/s10803-

018-3512-2

Wechsler, D. (2017). *Wechsler Individual Achievement Test – Third UK Edition (WIAT-III).* Pearson.

Weschler, D. (2008). *Weschler Adult Intelligence Scale version IV*. Pearson.

Whitby, M. (2018). Composite Mapping for Evaluation of Usability of Targeted Affordances 1 Breaking

Down Barriers to Usability of Inclusive Affordances for Neurodiverse Populations. In P.

Langdon, H. Dong, J. Lazar, & A. Heylighen (Eds.), *9th Cambridge Workshop on Universal*

Access and Assistive Technology. Cambridge University.