Assessment & Development

Matters

The difficulties of finding appropriate norms for assessing adults for dyslexia

Ian Smythe

T HE ASSESSMENT of dyslexia with adults has many challenges, including a shortage of appropriate tests. To help overcome this, some guidelines suggest using US tests to supplement the UK normed tests. However, this raises serious questions about validity, as well as the potential for institutional ageism. Whilst it is agreed that something should be used, it is important that the full implications are clearly understood.

Dyslexia assessment is problematic at the best of times. However, when assessing adults, there are additional difficulties, such as the lack of test material with appropriate UK norms. As a consequence, some guiding authorities are recommending using US normed tests. Whilst appearing to offer a viable alternative, the validity is, at best, questionable. However, rather than consider diverse implications (e.g. the impact of a 17 per cent Hispanic population on US norms), the intention here is to focus on tests where the US norm is available for older individuals (over 20), but UK norms are not available.

The legitimacy of an assessment process lies largely in the comparison of agreed criteria with an established norm, and identifying that somebody scores low on those norms would suggest that they may require special attention. This is true of most assessment processes in both health and education. In most medical areas, guidelines are well established, and embodied in widely used reference points such as DSM-5 and ICD-10. Those guidelines create the basis for that legitimacy, both of the assessment report produced and the person producing the document. However, in the case of dyslexia, the issue is more complex, largely due to a lack of a single, widely accepted reference source to create that unified legitimacy and the multi-dimensional nature of dyslexia.

Using a generic definition that dyslexia is 'a difficulty in the acquisition of fluent and accurate reading, writing and spelling' (Smythe, 2010, p.18), the assessment process only requires a verification of this information to confirm the diagnosis (N.B. All information over and above that would be considered a needs analysis, and not central to the identification). Thus there are two components: a) measurement of fluency and accuracy of the individual's reading, writing and spelling compared to the peer group, and b) establishing the difficulty in acquiring the skills, which can be through looking backwards at the educational history of the individual (e.g. provision of appropriate education, missed school for medical reasons, etc.) or forward looking as a response to (future) intervention, thereby including the time (difficulty in acquisition) dimension. It is the former, and in particular identification in comparison of adults to their peers, that is the focus of this article.

If resources for supporting dyslexic individuals are limited, and are distributed based upon the degree of difficulties experienced by the individual, it is necessary to compare the individual to others within the peer group to establish the legitimacy of resource allocation. This cannot be achieved without reference to a measure of that skill within the peer group.

Having moved away from the patronising reporting of grade level for adults (using child normed tests) to standard scores within adult normed tests, there is a need to have access to appropriate test materials, with norms available from 5 to 95. Thanks to the guidelines, assessors have the apparent freedom to 'borrow' from suites of US tests, even though UK norms are not available. However it is important to understand how much equivalence there is, and the degree to which a given test may be deemed appropriate. For example, the SASC guidelines (SpLD Assessment Standards Committee, 2013) for usable tests include TOWRE-2, Woodcock Reading Mastery and the Wide Range Achievement Test 4 (WRAT 4), none of which currently have UK norms. But what happens when somebody is reassessed using a UK norm instead of a US norm. For the purposes of illustration, let us consider the two possible scenarios, which do not relate to any specific test or sub-test:

■ The US population have greater skills that the UK population.
■ The UK population have greater skills than the US population.

As shown in Figure 1, if the UK norm is higher (A), an individual may not get support when using US norms. For example, if they scored 20 on a US normed test, which showed them to be 1.3 SD below the norm in the US, this would not meet the criteria to trigger support (assuming 1.5SD or more below the mean). However, if they had achieved this score on the same test with UK norms, they may be 1.6 SD below the mean, thus triggering support. Conversely, if the US norm is higher (B), support may be given where it would not be given if UK norms were used.
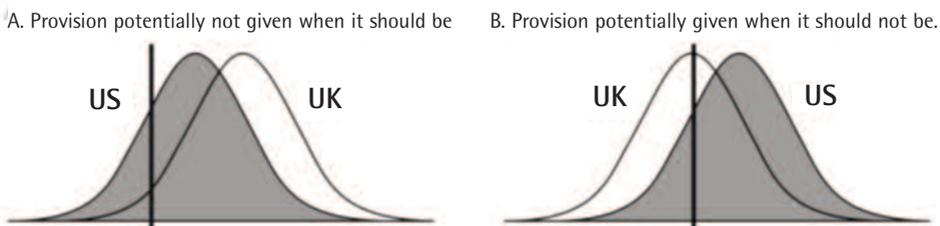
A. Provision potentially not given when it should be   B. Provision potentially given when it should not be.

Figure 1: Potential impact of using US vs UK norms

The problem is that there is no unequivocal evidence to confirm which is higher. The one (arguably) independent comparison is the results of Programme for International Student Assessment (PISA) scores. However, considerable doubt has been raised over the validity of the results, usually expressed as a ranking (Wuttke, 2007). For example, in 2009, the US was ranked above the UK in reading, but in 2012 this situation was reported as reversed. This is probably not so much about short term educational reform, but the precision of the statistics, where researchers suggest that the position of any country in the table is ±6 places.

Whilst it may be suggested that the assessment process usually involves many tests and, therefore, the impact of a single test will not have a significant impact, there are problems with this argument. Firstly, it assumes the impact is small and only one foreign norm is being used, and secondly that this single test would not affect the overall outcome of the assessment. However, if it does not contribute to the outcome of the assessment, why include it? That one test may have the casting vote which tips the balance towards providing support or not. Furthermore, if the final decision is being made with information outside the testing process (i.e. a decision as to whether to make allowances for diverse reasons), this undermines the process itself.

The second issue is with regards to the norming population and the validity of the comparison. The Wide Range Achievement Test 4 (WRAT 4) standardisation was based on a representative US national sample of approximately 3000 individuals ranging in age from 5 to 94 years, with 100–150 participants within each 'age group' (Wilkinson & Robertson, 2006). Similarly if we look at norming for the Wide Range Intelligence Test (WRIT), we find that only 2285 individuals were involved in the standardisation, ranging from 4 to 85 years old, with 87-125 participants within each 'age band' (Glutting et al., 2000). The European Federation of Psychologists' Associations' (EFPA) model for reviewing psychological tests suggests that a sample of 150 per age band is necessary (EFPA, 2009). Both these examples suggest that these US tests need careful scrutiny before being recommended as usable within a UK population.

Conclusions

There is no doubt that one has to be pragmatic regarding assessment practices. However, there is another issue, within the process – institutional ageism within the assessment process. For example, if a 20-year-old scores equally poorly as a 60-year-old on a reading comprehension test, the 20-year-old's score may be significantly lower than expected for his/her age group, whilst the 60-year-old's score may not be significantly below that expected for his/her age group. Could this lead to the younger student getting support and not the older student? Is that morally right? Indeed, is the older dyslexic student being penalised (through age discrimination) due to having greater life experiences that help hide (and overcome) some of the difficulties? Does this mean we should question who should be regarded as the peer group?

Knowing the vagaries of the legal system, it would be inappropriate to suggest where the courts would determine the legal responsibility lies within these thorny issues. But maybe it is another good reason to check the fine print on one's professional indemnity insurance.

The author

Dr Ian Smythe is a researcher in the field of reading and writing and director of Do-IT Solutions Ltd.

References

EFPA (2009). *EFPA review model for the description and evaluation of psychological tests.* Retrieved 9 October 2014 from www.psychtesting.org.uk/test-registration-and-test-reviews/test-reviews-and-test-registration_home.cfm

Glutting, J., Adams, A. & Sheslow, D. (2000). *Wide range intelligence test: Manual.* Wilmington, DE: Wide Range Inc.

Smythe, I. (2010). *Dyslexia in the digital age.* Chichester: Wiley.

SpLD Assessment Standards Committee (2013). *Suitable tests for the assessment of specific learning difficulties in higher education* (revised September 2013). Retrieved 19 September 2014 from www.sasc.org.uk/SASCDocuments/REVISED%20guidelines-September%202013%20v2.pdf

Wilkinson, G.S. & Robertson, G.J. (2006). *Wide range achievement test 4: Professional manual.* Lutz, FL: Psychological Assessment Resources Inc.

Wuttke, J. (2007). Uncertainties and bias in PISA. In S.T. Hopmann, G. Brinek & M. Retzl (Eds.). *PISA according to PISA.* Retrieved 29 September 2014 from www.univie.ac.at/pisaaccordingtopisa/pisazufolgepisa.pdf