

The dangers of using “foreign” norms for dyslexia assessment in adults

Ian Smythe

In this article, the vexing question of acceptability of using US normed tests in a UK context is addressed. Whilst much debated, most responses seem to ignore legal and moral considerations as well as that of the dangers of combining diverse cognitive domains into one metric.

Introduction

Dyslexia assessment is problematic at the best of times. However, when assessing adults there is the added difficulty of a lack of tests 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 17% Hispanic population on US norms), the intent 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 references points such as DSM-V and ICD-10. Those guidelines create the basis for that legitimacy, both of the document 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), the assessment process only requires a verification of this information to confirm the diagnosis. (N.B. All other information over and above that would be considered a needs analysis, and not central to the identification, for which only those areas that are noted as characteristics in the definition are required.) Thus there are two components: a) measurement of fluency and accuracy of the 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 education 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 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 are 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.

This leads to two important questions:

Are US norms appropriate for use in the UK?

What are the legal (and moral) implications for using US normed tests?

Are US norms appropriate for use in the UK?

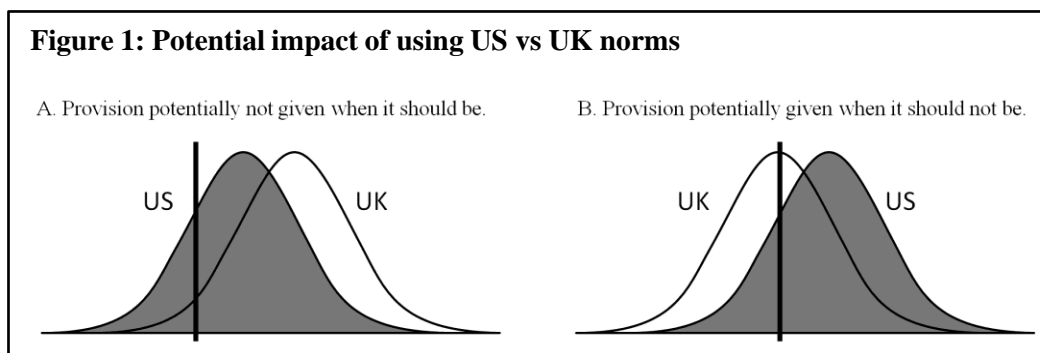
The important question is how much equivalence is there, and the degree to which a given test is deemed appropriate. For example, the SASC guidelines (SASC, 2013) for usable tests include TOWRE-2, Woodcock Reading Mastery and WRAT 4, none of which currently have UK norms. But what happens when somebody is reassessed using a UK norm instead of a US norm. (N.B. For the purposes of illustration, consider this to be a “thought game” rather than highlighting any specific test or sub-test.)

Despite the lack of verification of the merits of each country, let us consider the two possible scenarios:

The US have greater skills than the UK population

The UK has greater skills than the US population.

As shown in Figure 1, if the UK norm is higher (A), they may not get support when using US norms. For example, say they scored 20 on a US normed test, which showed them to be 1.3 SD below the norm in the US, not enough to trigger support at 1.5SD. But if they had used the same test with UK norms, they may be 1.6 SD below, thus triggering support. Conversely, if the US norm is higher, support may be given where it would not be given if UK norms were used.



The problem is that there is no unequivocal evidence as to which is higher, and few researchers are interested in this subject since data set size (and consequential funding) will need to be huge to be robust. A small number of test developers offer both US and UK norms, but it is difficult to interpret marketing material to identify if the two versions exist because there is a difference between them, or because they want to suggest it has been made appropriate for the UK. The one (arguably) independent comparison is the results of international PISA scores (Wuttke, 2007). However, considerable doubt has been raised over the validity of the results usually expressed as a ranking. 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.

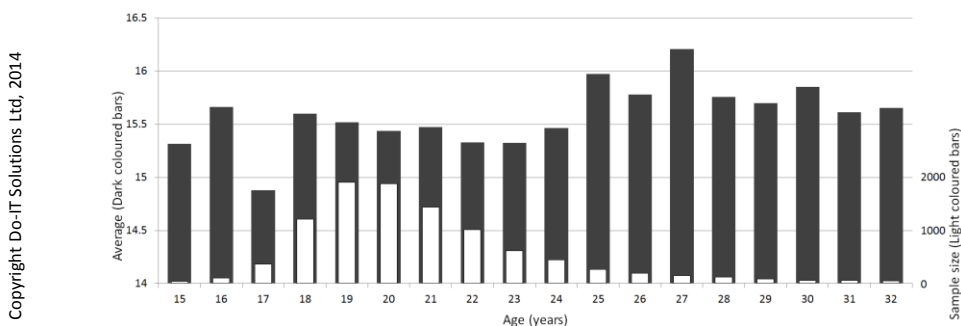
Whilst it may be suggested that the assessment process usually involves many items and therefore the impact of a single item will not have a significant impact, there are problems with this argument. Firstly, it assumes the impact is small and only one “foreign” norm is being used, and secondly that this single item would not hold the casting vote, which may occur in some circumstances. (Put another way, if it does not contribute to the decision, why include it?) If this is not the case, then the decisions are being made with information outside the testing process, which would seem to undermine the process itself.

What are the legal (and moral) implications for using US normed tests?

The reason for using the US test is due to a lack of norms for ages over 18 years old. But this itself is problematic since it leads to a possible scenario whereby two individuals can have the same degree of difficulty, differing only in age, and one may get support and the other may not. The issues are complex (and outside this article), but recent research suggests greater consideration needs to be given to the nature of assessment and the type of support given. For example, research on memory (Sander, Lindenberger and Werkle-Bergner, 2012) suggests differential decay rates of brain structures used for components of memory means that the precise nature of the test (e.g. Reading Comprehension) will significantly impact upon results. That said, some research (see Figure 2, but also diverse cognitive measures) questions the need to have norms for each year over 16 years old.

Figure 2: Reading Comprehension average versus age

Below is a graph of the reading comprehension scores from 10,000 students from Further Education and Training Colleges in Johannesburg. Variation of scores with age are less than the precision of the test, suggesting that for this cohort, age differentiation is not required.



Again using an example, if a 20 year old scores equally poorly as a 60 year old on a comprehension test, the norms that expect lower scores due to poorer memory may mean the young student gets support and not the older student. Is that morally right? Indeed, is the older dyslexic student being penalised (age discrimination) due to having greater life experiences that help hide (and overcome) some of the difficulties.

Conclusions

In the end, are we asking the right question? There is no doubt one has to be pragmatic. But at the same time as providing pragmatic resolutions, it is important to ensure all parties are clear about the implications. Knowing the vagaries of the legal system, it would be inappropriate to suggest where the courts would determine the legal responsibility lies within this thorny issue. But maybe it is another good reason to check the fine print on one's professional indemnity insurance.

About the author

Dr Ian Smythe is a researcher in the field of reading and writing as well as director of Do-IT Solutions Ltd, where he has access to significant data sets that help address such questions as posed here.

References

Wuttke J. (2007) Uncertainties and Bias in PISA. In Hopmann ST, Gertrude Brinek G. and Retzl M.(Eds.) PISA According to PISA. Lit Verlag GmbH. Vienna. Web access:

<http://www.univie.ac.at/pisaaccordingtopisa/pisazufolgepisa.pdf>

Sander M.C., Lindenberger U. and Werkle-Bergner M. (2012) Lifespan age differences in working memory: A two-component framework. *Neuroscience and Biobehavioral Reviews*. P2007-2033.

SASC (2013) Suitable Tests for the Assessment of Specific Learning Difficulties in Higher Education (Revised September 2013) Web access: www.sasc.org.uk. Access date 19/09/2014

Smythe I (2010) *Dyslexia in the Digital Age*. Wiley. Chichester.