The British
Psychological Society
Psychological Testing Centre

Assessment & Development Matters

Vol. 6 No. 3
Autumn 2014

ISSN 2040-4069

Inconsistency in dyslexia assessment

Ian Smythe

INTER-RATER RELIABILITY refers to the degree of agreement between two or more human evaluators, and is rarely carried out in the field of dyslexia for a variety of reasons, including cost, time, stress to the assessee and the implicit dangers of such analyses. Consequently research and commentary in the field is scarce. However, some research in the area of artificial neural networks provides the opportunity to reflect on some of the issues. Dyslexia is a difficulty in the acquisition of fluent and accurate reading and writing skills (Smythe, 2010). Whilst vociferous debates proliferate around the validity and usefulness of the construct, few would argue the existence of a significant number of individuals who have neurological differences that impact on cognitive processing sufficiently to cause a difficulty in acquiring good literacy skills. However, there has been much voicing of concern around assessment inconsistency, though few have committed that in writing.

As Brown and Loosemore (1995) noted, you can often understand if a model is robust if you try to computerise it. The difficulty is that the relationships between cognitive processes (and the symptoms they affect) are rarely linear and there are multiple data points. Therefore, any attempt to use traditional statistical interpretation is at best problematic, at worst doomed to failure. However, in many situations, if we cannot see clearly the relationship between the multitude of components that we are trying to analyse, an experienced person can call on ‘intuition’ to arrive at an answer. Recent artificial intelligence techniques, such as artificial neural networks, can identify underlying hidden patterns in complex data sets even when they are non-linear. However, this is a ‘black box’ method, which means that though the outcomes are valid, there is little insight into the underlying structure. Therefore, it may be argued that the result is analogous to intuition. Turing (cited in Hodges, 1997) suggested that ‘The activity of the intuition consists of making spontaneous judgments which are not the result of conscious trains of reasoning. These judgments are often but by no means invariably correct (leaving aside the question what is meant by “correct”).’ Often it is possible to find some other way of verifying the correctness of an intuitive judgment. Dyslexia assessment, at times, may seem to lend itself more to intuition than cognitive science, particularly in borderline cases.

Using artificial intelligence as arbitration

Dyslexia assessment involves complex, multi-dimensional, multi-criteria cognitive processes that need to be evaluated to produce a one-dimensional outcome – dyslexic or not – albeit on a continuum of difficulties. There have been various attempts by researchers over the past 40 years to use artificial intelligence to automate and validate the process, the most promising of which are fuzzy logic (where allowance is made for imprecision), expert systems (where experts agree the outcomes from given scenarios) and artificial neural networks. However, the inconsistency of results has been blamed on the computer techniques, with little consideration of potential inconsistencies in the (independent) diagnostic process.

Artificial neural networks (ANNs) have the capability to handle the multiple data points, derive criteria and handle non-linear cognitive relationships that occur in dyslexia assessment, and appear to offer the greatest promise. Various research groups (Jain et al., 2009; Kohli & Prasad, 2010; Upandhyay et al., 2013; Wu et al., 2006) used cognitive assessment to train the ANNs to predict assessment outcome, comparing results with independent assessments. All researchers questioned the ability of their ANNs to accurately predict at a consistency level greater than 85 per cent, but nobody questioned the validity of the assessments (N.B. All research teams were led by computer scientists, and rarely had support of education and/or psychology specialists.) However, Wu et al. (2006) noted 'that certain students manually diagnosed as LD (learning disabled) are always classified as non-LD, no matter what technique is used' (p.8826). They also observed that 'if proven feasible, computer-aided diagnosis can not only save time and manpower, but also have the advantage of eliminating possible human bias' (p.8826).

Research has shown that where there is the strong inter-rater reliability required to train the algorithms (e.g. fruit classification using image analysis), ANNs can predict the outcomes given enough data points. Where classification contains subjective components, such as in the field of dyslexia, consistency is prone to fall and computer-based predictions are more difficult to validate. But this also begs the question that if the computer suggests one outcome and a human assessor suggests another, who is right?

Potential sources of errors

There are many potential sources of error, but for the sake of brevity, consider that the core of many assessments includes executive functioning (including working memory) and reading skills. However, each is poorly defined and tests that claim to measure those constructs can produce wildly different results. For example, there are many suggested ways to evaluate executive functioning because there is little agreement on the theoretical model, and reading skills measures have limited correlation as they measure different constructs. Furthermore, the test usually relates to a specific skill, and not to dyslexia diagnosis. Therefore the assessor will set their own criteria and interpretation to assist 'proving' (or disproving) the dyslexia. Furthermore, there are the 'human traits' in interpretation, which means the outcomes may be a function of assessor history, knowledge of the person outside the test protocol, and the degree to which they have updated their skills in line with recent research and debate. Potential conflicts of interest may also be a factor. Whilst continuing professional development has made significant strides in this field in recent years, the inconsistencies are still greater than would be acceptable in other professions.

Outcome and conclusion

Clearly the arguments are more complex than room permits here. This, however, is not simply a technology or philosophical argument. South Africa has an emerging disability support system where funding is now available for dyslexia support. However, the assessment system lacks the maturity of the UK system. A lack of dyslexia history, clear criteria and trained professionals as well as assessment tools that can deal with 11 national languages means that distribution of available funding will be sporadic. Whilst there is a long way to go before a robust system is in place, it would seem logical that some form of online assessment validation process could be implemented at the start in order to provide consistency in funding allocation. That is, the assessment could be submitted for review within a number of widely accepted models, and if the assessment outcome fails all models, then greater justification for the diagnosis is required before it can be accepted.

This approach could combine ANNs with other artificial intelligence techniques such as fuzzy logic and expert systems (possibly augmented by Delphi and Analytic Network Process methods) to minimise the inconsistencies that otherwise will ensue, and ensure that the allocation of funds to dyslexic individuals is based on consistent test analyses, and not on 'intuition'. Interestingly, this is little more than an updated version of a system implemented by the California Community Colleges designed to 'reduce or eliminate the inequities, inconsistencies and biases that characterised previous eligibility models' (p. 1) originally implemented more than 25 years ago.

The author

Dr Ian Smythe is a researcher in the field of reading and writing. He is particularly interested in the use of artificial intelligence to help identify the hidden talents of those with learning difficulties.

References

Brown, G.D.A. & Loosemore, R.P.W. (1995). Computational approaches to normal and impaired spelling. In G.D.A. Brown & N.C. Ellis (Eds.) *Handbook of spelling: Theory, process and application* (pp. 319-36). Chichester: John Wiley.

California Community Colleges (2002). *Learning disabilities eligibility model.* Sacramento, CA: California Community Colleges Chancellor's Office.

Hodges, A. (1997). *Alan Turing: One of the great philosophers.* Retrieved 25 June 2014 from www.turing.org.uk/publications/ex4.html.

Jain, K., Manghirmalani, P., Dongardive, J. & Abraham, S. (2009) Computational diagnosis of learning disability. *International Journal of Recent Trends in Engineering, 2*(3), 64–66.

Kohli, M. & Prasad T.V. (2010). Identifying dyslexic students by using artificial neural networks. *Proceedings of the World Congress on Engineering*. Retrieved 1 July 2014 from www.iaeng.org/publication/WCE2010/WCE2010_pp118-121.pdf.

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

Upandhyay, A., Singh, S.K., Turkar, V. & Singh, O. (2013). Classification of learning disabled students using artificial neural network. *Proceedings of the ITR International Conference*, Bhubaneswar, India, pp.1-4.

Wu, T.K., Huang, S.C. & Meng, Y.-R. (2006). Identifying and diagnosing students with learning disabilities using ANN and SVM. *Proceedings of the IEEE International Joint Conference on Neural Networks*, 8820–8827.