An analysis of differential item functioning by gender in the Learning Disability Screening Questionnaire (LDSQ)

Aja Louise Murray [a,*], Tom Booth [a], Karen McKenzie [b]

[a] Department of Psychology, University of Edinburgh, UK
[b] Department of Psychology, University of Northumbria, UK

ABSTRACT

The Learning Disability Screening Questionnaire (LDSQ; McKenzie & Paxton, 2006) was developed as a brief screen for intellectual disability. Although several previous studies have evaluated the LDSQ with respect to its utility as a clinical and research tool, no studies have considered the fairness of the test across males and females. In the current study we, therefore, used a multi-group item response theory approach to assess differential item functioning across gender in a sample of 211 males and 132 females assessed in clinical and forensic settings. Although the test did not show evidence of differential item functioning by gender, it was necessary to exclude one item due to estimation problems and to combine two very highly related items (concerning reading and writing ability) into a single literacy item Thus, in addition to being generally supportive of the utility of the LDSQ, our results also highlight possible areas of weakness in the tool and suggest possible amendments that could be made to test content to improve the test in future revisions.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Intellectual disability (ID) is defined with respect to three criteria: significant impairment in intellectual functioning, significant impairment in adaptive functioning, and onset before adulthood (British Psychological Society, 2000). From a societal perspective, ID is an important disorder affecting approximately 10 in 1000 individuals and involving the utilisation of large amounts of public health resources (Maulik, Mascarenhas, Mathers, Dua, & Saxena, 2011). Accurately identifying individuals with ID is an important psychometric challenge. Correct identification of individuals with an ID facilitates the provision of appropriate services and resources to individuals in need of this additional support. However, false identifications waste resources and potentially expose the wrongly diagnosed individual to social stigma. Given the potential consequences of mis-classifying individuals as affected versus unaffected by ID, there are stringent guidelines regarding the assessment of the disorder. An assessment should be conducted by a trained professional with intellectual deficits measured using a standardised assessment such as the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV: Wechsler, 2008) in adults. It is also recommended that adaptive functioning is also assessed using a standardised assessment such as the Adaptive Behaviour Assessment System – Second Edition (ABAS-II: Harrison & Oakland, 2003) or Vineland Adaptive Behaviour Scales – Second Edition (Vineland II: Sparrow, Balla, & Cicchetti, 2005). The necessary rigour with which individuals are assessed makes the process of assessment time and resource intensive as well as demanding and potentially stressful for the

---

individual being assessed. For resource stretched psychological services and individuals with a potential ID alike, it is, therefore, desirable to target full diagnostic assessments only to those individuals with a genuine risk of having an ID.

The Learning Disability Screening Questionnaire (LDSQ: McKenzie & Paxton, 2006) was developed with this goal. Note that while the LDSQ measures 'intellectual disability' it was developed at a time when this construct was referred to by the label of 'learning disability' in the UK. This was prior to the adoption of the term 'intellectual disability' (see Schalock, Luckasson, & Shogren, 2007 for a discussion regarding changing terminologies). The LDSQ is a brief screening tool for identifying ID that can be used by a range of professionals without the need for extensive training. It provides a quick estimate of whether an individual is likely to have an ID or not and, therefore, aids in identifying those individuals who should undergo full assessment for ID. The content of the scale refers to key areas of intellectual and adaptive functioning that are associated with the presence of ID. Across seven dichotomously scored items the areas of practical living, occupational, and academic skills are assessed. Abbreviated item contents are provided in Table 1.

The utility of the LDSQ as a clinical and research tool has been evaluated in several previous studies. In terms of its utility as a screening tool, an initial validation study in a clinical sample reported sensitivity and specificity values of 91% and 87% respectively at the cut-off point specified for suspected ID in a community sample (Paxton, McKenzie, & Murray, 2008). Similar discriminative ability was later reported in a forensic sample (sensitivity = 82% and specificity = 88%; McKenzie, Michie, Murray, & Hales, 2012). A subsequent study evaluated the ability of the scale to accurately classify individuals in terms of the British Psychological Society severity categories of 'no ID', 'significant impairment' (IQ between 55 and 69) and 'severe impairment' (IQ less than 55) (Murray & McKenzie, 2014). This study found that, while the performance of the scale in making this 3 category distinction was statistically significant, it was not clinically significant in the sense of providing sufficient certainty of classification to justify a formal clinical categorisation. However, Murray and McKenzie (2013) did identify some additional benefits to the scale in terms of clinical and research applications. They found that the 6 out of 7 of the items exhibiting invariant item ordering: a property which means that information about levels of severity is contained not only in overall scores, but also in individual item responses. For example, if an individual is unable to tell the time (an easy item), it is possible to infer that their impairment is relatively severe and that it is unlikely that they can read or write (harder items), even if this information is not available.

In terms of applications of the LDSQ, it has been used in clinical and criminal justice services across the UK, in Europe, Japan, and Australia. It was independently piloted by the Department of Health for use in a range of criminal justice services and is currently being used by a number of probation, prison and police services, as well as ID services in the UK.

To date, however, no studies have examined the important issue of whether the LDSQ has equivalent measurement properties in males and females i.e. that it does not show test bias with respect to gender.

There are broadly two forms of test bias that can be defined: prediction invariance and measurement variance. Prediction invariance refers to when the relation between test scores and some criterion is equal across groups, whereas measurement invariance refers to when the relation between the test score and the latent attribute measured by the test is equal across groups. Borsboom, Romeijn, and Wicherts (2008) demonstrated that under most circumstances when one holds this implies violation of the other. Thus, a choice must be made between conceptualising test bias in terms of measurement or predictive invariance. The authors provide several arguments for preferring the measurement invariance approach, an important one being that only a measurement invariant test implies that individuals with the same latent trait level should have the same expected test score.

More specifically, measurement invariance can be defined as the conditional distribution of observed scores given the underlying latent variable of interest being independent of group membership (Kim & Yoon, 2011):

$$f(X|W, G) = f(X|W)$$

where X represents the observed score, W represents the underlying latent variable, and G represents group membership. The consequences of a lack of invariance in the context of research studies and selection have been widely discussed (e.g. Meredith & Teresi, 2006).

In the context of measuring a trait such as ID, the absence of measurement invariance can result in the test being biased against one of the groups, for example, with males having higher expected scores than females for the same standing on the latent trait. Thus, the test could systematically over-estimate the level of impairment of one gender relative to the other. It is

Table 1
LDSQ items and proportion of endorsement in males and females.

Item	Male		Female	
	N	Proportion endorsed	N	Proportion endorsed
Time	133	0.65	82	0.60
Read	180	0.42	118	0.47
Write	171	0.36	118	0.46
Literacy (read + write)	211	0.50	132	0.58
Independent living	199	0.23	132	0.34
Employment	197	0.21	132	0.22
Previous contact	191	0.32	129	0.33
School	189	0.24	131	0.29