Estimating the level of functional ability of children identified as likely to have an intellectual disability

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

[a] Centre for Cognitive Ageing & Cognitive Epidemiology, Department of Psychology, 7 George Square, University of Edinburgh, Edinburgh EH8 9JZ, Scotland, UK
[b] Department of Clinical Psychology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland, UK
[c] Psychological Services, NHS Borders, Andrew Lang Unit, Viewfield Lane, Selkirk TD7 4LJ, Scotland, UK

ABSTRACT

Screening tools can provide an indication of whether a child may have an intellectual disability (ID). Item response theory (IRT) analyses can be used to assess whether the statistical properties of the tools are such that their utility extends beyond their use as a screen for ID. We used non-parametric IRT scaling analyses to investigate whether the Child and Adolescent Intellectual Disability Screening Questionnaire (CAIDS-Q) possessed the statistical properties that would suggest its use could be extended to estimate levels of functional ability and to estimate which (if any) features associated with intellectual impairment are consistently indicative of lower or higher levels of functional ability. The validity of the two proposed applications was assessed by evaluating whether the CAIDS-Q conformed to the properties of the Monotone Homogeneity Model (MHM), characterised by uni-dimensionality, local independence and latent monotonicity and the Double Monotone Model (DMM), characterised by the assumptions of the MHM and, in addition, of non-intersecting item response functions. We analysed these models using CAIDS-Q data from 319 people referred to child clinical services. Of these, 148 had a diagnosis of ID. The CAIDS-Q was found to conform to the properties of the MHM but not the DMM. In practice, this means that the CAIDS-Q total scores can be used to quickly estimate the level of a person's functional ability. However, items of the CAIDS-Q did not show invariant item ordering, precluding the use of individual items in isolation as accurate indices of a person's level of functional ability.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The early recognition of intellectual disability (ID) is important to ensure appropriate support is provided to maximise the child's development (Guralnick, 2005); to facilitate access to resources (Goodman & Linn, 2003); to inform differential diagnosis; and to highlight situations where genetic testing or counselling may be required (American Academy of Pediatrics, Committee on Children with Disabilities, 2001). Many children, however, may not be identified as having ID until they are teenagers (Simonoff et al., 2006) or even young adults (Hamilton, 2006). One potential reason for this is that diagnosing ID can be complex and time consuming (Ryan, Glass, & Brown, 2007). The assessment of intellectual functioning,

in particular, can cause delays in the diagnostic process because it requires the use of a standardised and validated assessment that is individually administered by an appropriately qualified applied psychologist (British Psychological Society [BPS], 2000).

This has led to the recognition that screening tools may offer a pragmatic solution in circumstances where there is a desire to reduce waiting times and channel referrals appropriately (BPS 2003) by having a quick method of identifying those individuals who should undergo full assessment. Screening tools may also be used by researchers (see Charman et al., 2007) to identify particular populations of interest, where the need to assess a large number of individuals may make the use of full diagnostic assessments unfeasible. Screening tools may also be used where there is the need for an early indicator of the potential support needs of children, for example in educational settings (Sonnander, 2000), while waiting for full diagnostic assessment to take place. Under all of these circumstances, it would be clinically useful if screening tools could provide more information about the potential support needs of the child, to allow service planning to begin at an early stage.

Screening tools, however, tend to use a cut-off score that results in a dichotomous classification of either ID or non-ID, reflecting their aim of identifying individuals who may have ID and who, therefore, should undergo further assessment (McKenzie & Megson, 2012). If, however, a screening tool was able to give an indication of the extent of the functional ability of a child, or be used to indicate what abilities are expected to develop ahead of others, then it may have additional clinical and research benefits.

Item Response Theory (IRT) analysis can be used to evaluate whether a clinical scale possesses the statistical properties such that it can validly be used in more extended ways in clinical settings (Reise & Waller, 2009). For example, while ROC analysis can be used to assess the classification accuracy of a screen for ID (and thus focuses on the range of the trait close to the diagnostic threshold), IRT based techniques focus more explicitly on how the test performs across the full range of the latent trait. Similarly, factor analyses can be useful in assessing the dimensionality of a scale and how well particular items measure identified dimensions, however, there is generally not an explicit focus on how items perform at different locations on these dimensions.

In general, IRT approaches offer many advantages and analysis options that are not readily available in these more traditional approaches to test development (Embretson & Reise, 2000). Meijer and Baneke (2004) have argued for the utility of non-parametric IRT models, in particular for the analysis of psychopathology scales. Unlike parametric IRT models, non-parametric models do not impose a specific structure on the relation between item responses and the underlying latent trait (the 'item response function' or IRF). This is an advantage when assumptions about the form of the IRF, for example the logistic function, are unrealistic for empirical data. Meijer and Baneke (2004) note that non-parametric IRT models can provide useful information about the performance of items and scales without the need to make these assumptions. In addition, samples in clinical studies tend to be of only modest size due to the relative infrequency of clinical disorders in the population and/or recruitment difficulties. Non-parametric IRT models generally have smaller sample size requirements than parametric IRT models.

Mokken scaling is a non-parametric IRT method that can be used to investigate important and clinically useful properties of scales (e.g., see Stochl, Jones, & Croudace, 2012). First, it can be investigated whether a scale conforms to the properties of the monotone homogeneity model (MHM) that is characterised by the assumptions of uni-dimensionality, local independence and latent monotonicity. When the assumptions of MHM hold, it is possible to infer stochastic ordering on the latent trait, that is, that higher test scores are probabilistic indications of a higher level on that trait. Although scale scores are frequently assumed to possess this property, it is important to explicitly test this assumption (Meijer & Baneke, 2004). In practical terms, evidence that MHM holds for a given scale provides some justification for the use of the scale in clinical practice for tasks that require an ordering of individuals based on severity of the trait of interest (e.g., referral for treatment).

It is also possible to investigate whether a scale conforms to the properties of the double monotonicity model (DMM). This is characterised by the assumptions of the MHM, plus the additional assumption of non-intersection of item response functions. When the DMM holds, items form a consistent hierarchy and both items and people can be characterised by their position on a continuum defined by levels or severity of the latent trait (e.g., 'level of functional ability'). That is, items located at higher levels of the latent trait (e.g., those requiring higher levels of functional ability) tend only to be endorsed if items located at lower levels of the latent trait (e.g., those requiring only a moderate level of functional ability) have also been endorsed. This means that items indicative of a more severe impairment will consistently be endorsed ahead of those indicative of a less severe level of impairment. As such, if DMM holds in a given scale, clinicians may be able to gain important information on the symptomology of a given disorder by the location of specific items on the severity continuum. Furthermore, investigating whether the MHM and DMM hold for a particular scale yields important information on item performance within the scale, and thus allows researchers and test developers to improve the utility of clinical scales.

Several authors have recently championed Mokken scaling in clinical measures as a means of enhancing their clinical utility (e.g., Watson et al., 2012). For example, Mokken scaling analysis has proven useful in disability research where estimating the severity or predicting the progression of difficulties is a key consideration in understanding and treating a disorder (Kingston et al., 2012; Watson et al., 2012). Similarly, Murray and McKenzie (2013) applied Mokken scaling analysis to an adult ID screening tool: the Learning Disability Screening Questionnaire (LDSQ; McKenzie & Paxton, 2006) and found that the scale conformed to DMM. Thus, information on responses to single items (not just total scores on the whole scale) are informative about a person's likely functional difficulties. That is, even if total scale scores were not available for an

individual, it would be possible to estimate the severity of their impairment based on their responses to individual items because these were known to form a consistent hierarchy. This may be especially beneficial in a clinical context where individuals with ID may have difficulties in comprehension and communication, leading to ambiguities with regards to their responses to certain items.

Given the importance of early identification of the potential support needs of children with ID, the aim of the present study was to apply Mokken scaling analysis to a child screening tool, to determine if it, and the individual items comprising it, had the statistical properties that would allow it to be informative about an individual's functional ability. The chosen screening tool was the child version of the LDSQ: the Child and Adolescent Intellectual Disability Screening Questionnaire (CAIDS-Q; McKenzie, Paxton, Murray, Milanesi, & Murray, 2012). The CAIDS-Q was considered appropriate for this investigation for a number of reasons: it was developed specifically as a screen for ID; it was designed to have a very short administration and scoring time; it does not require the user to have a particular professional qualification or training; and a series of studies have found it to have good psychometric properties (see Section 2.2). In addition, two recent reviews of screening tools that have been used specifically to screen for ID in children (McKenzie & Megson, 2012) and those that have screened for a range of disabilities, including ID (Maulik & Darmstadt, 2007) indicated that none of the reviewed tools displayed sufficiently robust psychometric properties to recommend their use for specific screening for ID in children. The CAIDS-Q had not been developed at the time of these reviews.

In the present study we, therefore, assessed whether the CAIDS-Q possessed the statistical properties that would allow it to be applied in ways that extend beyond its original application as a screening tool. Specifically, we investigated whether the scale conformed to the MHM and the DMM. This indicated whether scale total scores and individual items respectively could be used to estimate levels of functional ability.

2. Method

2.1. Participants

We conducted analyses on data from the participant sample used in the development and validation of the CAIDS-Q (McKenzie, Paxton, Murray, et al., 2012), plus some additional data gathered since the first validation study was published (total $N = 319$) for which ethical approval had previously been given by the relevant Caldicott Guardians. In Scotland, it is the role of the Caldicott Guardian, on behalf of each health board area, to ensure that patient data are used appropriately for research and audit purposes, where it is not feasible to contact each individual to obtain consent to use existing data. Pre-existing information for children who had been referred to ID/child and adolescent mental health (CAMH) services in four National Health Service areas in Scotland was used for the analyses. Data were collected, where available, in relation to the CAIDS-Q items, Full Scale IQ, whether the individual had been given a diagnosis of ID or not, age at assessment and gender. The reason for referral for each individual was not noted, but all individuals had an assessment of intellectual and adaptive functioning and developmental history carried out by the service as part of the assessment process.

Table 1 contains descriptive information on the age and FSIQ of the sample by gender and ID diagnosis. Of these participants, 148 (male = 88; female = 58; 3 gender unknown) had a diagnosis of ID. Diagnosis of ID was determined, as noted in the case-notes, by the independent clinician based on the diagnostic criteria for ID.

We did not stratify our sample by age because this would have reduced sample size to levels not appropriate for Mokken scaling analysis. We conducted analyses on the 314 cases that had complete data on the CAIDS-Q, which resulted in the deletion of 5 cases from the total sample. Comprehensive descriptions of these samples can be found in McKenzie, Paxton, Murray, et al. (2012) and McKenzie, Paxton, Michie, et al. (2012).

2.2. Measures

The CAIDS-Q is a seven item instrument developed for the purpose of identifying individuals at risk of having ID. Items are scored dichotomously and refer to the capabilities of the child in several domains in which individuals with ID may experience difficulties, including literacy, self-care, relationships, and previous and current support by specialist services. A percentage score is calculated, with higher scores on the CAIDS-Q being indicative of a higher level of ability. Scores below a certain threshold, which varies depending on a child's age, suggest that the child may have an ID. Thus, based on age group-specific cut-off scores, the CAIDS-Q assigns a child as being likely to have an ID, which suggests that they should undergo full assessment for ID, or as not being likely to have an ID.

Previous research has examined the psychometric properties of the scale in clinical (McKenzie, Paxton, Murray, et al., 2012) and forensic (McKenzie, Paxton, Michie, et al., 2012) samples and have supported the uni-dimensionality, reliability, criterion validity and predictive ability of the scale. Evaluations with clinical samples of children aged 6–7 years 11 months (McKenzie, Murray, & Murray, 2013) and young people aged 8–18 years (McKenzie, Paxton, Murray, et al., 2012) have found sensitivity and specificity values of 82% and 83% respectively for the younger group and 96% and 85% for the older group, at the relevant cut-off scores. It has also been found to correlate well with both intellectual (McKenzie, Paxton, Murray, et al., 2012) and adaptive functioning (McKenzie & Murray, in press) and to perform well when compared with a short form intellectual assessment (Mckenzie, Murray, Murray, & Murray, 2013). It has, not, however, yet been tested whether the CAIDS-Q items conform to the properties of the MHM and DMM.

Table 1
Sample descriptive statistics.

	Total sample ($N = 319$)		Males ($n = 202$)		Females ($n = 113$)	
	ID ($n = 148$)	No ID ($n = 171$)	ID ($n = 88$)	No ID ($n = 114$)	ID ($n = 58$)	No ID ($n = 55$)
Age mean (SD)	151.43 (33.08) $n = 145$	146.51 (33.28) $n = 166$	151.41 (35.57) $n = 86$	146.29 (33.60) $n = 111$	152.04 (29.71) $n = 57$	145.98 (33.15) $n = 53$
FSIQ mean (SD)	54.28 (9.59) $n = 146$	87.97 (16.21) $n = 162$	54.03 (9.16) $n = 87$	86.34 (13.89) $n = 108$	54.95 (10.24) $n = 57$	92.04 (19.74) $n = 52$

Note: n within cells denotes the amount of available data used for calculations.

Table 2
Explanation of Model Assumption and Clinical Implications for the CAIDS-Q.

Assumption	Explanation	Clinical implication
Uni-dimensionality	Responses to items are caused by a uni-dimensional latent trait.	The scale measures functional ability as a unitary construct.
Local independence	Items are independent conditional on the latent common factor.	Responses to items are not related to responses on other items except by virtue of the fact that they all measure the same underlying construct of functional ability.
Monotonicity	As the level of a latent trait increases, the probability of endorsing an item increases or stays the same.	The scale is an ordinal person scale, i.e., higher scale scores are suggestive of greater functional ability.
Non-intersection	The item response functions of different items do not cross one another.	The items of the scale form a consistent hierarchy across individuals. This means that individual items can be used to estimate a person's level of functional ability.

Notes: The first three assumptions characterise the MHM. The DMM is characterised by all four assumptions.

2.3. Mokken procedure

Mokken scaling analysis allows the testing of the assumptions of non-parametric item response models (IRT). The two most popular models are the general non-parametric IRT model: the MHM, and the more restricted DMM (Van der Ark, 2012). The MHM is defined by three assumptions: uni-dimensionality, local independence and latent monotonicity. The DMM is defined by these three assumptions, plus the assumption of non-intersection of item response functions. An explanation of these assumptions and how they are tested are provided in Table 2. When the assumptions of MHM hold, it is possible to infer stochastic ordering on the latent trait being measured by the scale, that is, that higher test score are probabilistic indications of a higher level on that trait. As the DMM is defined by the additional assumption of non-intersection of item response functions, when it holds it is also possible to infer invariant item ordering for dichotomous items (Van der Ark, 2012). For polytomous items, additional conditions must be met in order to infer invariant item ordering, however, here we analyse dichotomously scored items, making non-intersection of item response functions sufficient to infer invariant item ordering (Sijtsma, Meijer, & Van der Ark, 2011).

We conducted all analyses in the R package Mokken (Van der Ark, 2007, 2012). We first implemented the automated item selection procedure that selects items into scales using a hierarchical clustering algorithm. The algorithm selects items into a given scale beginning with the pair of items with the highest item pair scalability coefficients and proceeds until no more items remain that meet the criteria of a Mokken scale. Given that the CAIDS-Q is assumed to be uni-dimensional, we expected all seven items to be selected into a single scale.

Next, we obtained item scalability coefficients (H_i). These convey information about the relation between an item and the latent trait, reflecting item discrimination and degree of association with the latent trait. Items in the same Mokken scale should all have item scalability coefficients (H_i) larger than .3. We also examined total scale scalability coefficients (H), which express the strength of the scale as a whole. Mokken (1971) recommended the following criteria for judging the strength of Mokken scales: $H < .30$ is indicative of a weak scale; $.4 < H < .5$ is indicative of a moderate scale; and $H > .5$ is indicative of a strong scale.

We then proceeded to investigate the assumption of latent monotonicity. For dichotomous items, latent monotonicity implies manifest monotonicity, making it possible to infer the former from the latter. We checked for violations of manifest monotonicity, that is, instances in which the item response function does not increase with the level of the latent trait. As recommended by Van der Ark (2007), we ignored trivially small violations by considering violations that exceeded a certain minimum size only. For this purpose we adopted the default criterion of the package of a minimum violation size of .03.

Finally, we investigated the assumption of non-intersection of item response functions. We did this using the pmatrix method. This method examines the proportion of relative positive response, that is, $P(++)$ and proportion of relative negative response, that is, $P(--)$ item matrices. If non-intersection holds, $P(++)$ should be non-decreasing across all its rows and columns and $P(--)$ should be non-increasing across all its rows and columns. As in the investigation of manifest monotonicity, it is desirable to ignore trivially small violations of non-intersection. We, therefore, again adopted the programme default minimum size of violation of .03 as the threshold below which violations were not considered.

3. Results

The automated item selection procedure selected all items of the CAIDS-Q into the same Mokken scale, leaving no items unselected. The item scalability coefficients for items and item-pairs within this scale are provided in Table 3, together with the percentages of endorsement for each item.

The pmatrix method revealed a number of significant violations of the assumption of non-intersection of the item response function. These were spread across six of the seven items and could not be easily addressed by deletion of only a few items. Consistent with this, percentage endorsement of the items in the scale were, with the exception of the 'School Support' item, close in magnitude. The 'School Support' item was endorsed only 28% of the time whereas the other 6 items had percentage endorsements within a narrow range of 53–58%. Thus, items for the most part had similar locations along the latent trait continuum.

Table 3
Scalability coefficients.

Item	Item endorsement (%)	Item and item-pair scalability coefficients (SE)						
		1.	2.	3.	4.	5.	6.	7.
1. Read	58	.68 (.03)						
2. Friends	58	.39 (.06)	.42 (.05)					
3. Write	56	.96 (.02)	.44 (.06)	.68 (.03)				
4. Time	55	.82 (.05)	.38 (.06)	.77 (.05)	.65 (.03)			
5. Previous contact	53	.55 (.06)	.36 (.06)	.56 (.06)	.56 (.06)	.54 (.04)		
6. Shoe laces	53	.64 (.06)	.44 (.06)	.63 (.06)	.70 (.05)	.50 (.06)	.58 (.04)	
7. School support	28	.73 (.08)	.59 (.09)	.74 (.07)	.67 (.08)	.81 (.06)	.54 (.09)	.68 (.06)
Scale total H	.60 (.03)							

Notes: Below the diagonal are item-pair scalability coefficients. Item scalability coefficients are on the diagonal. Items are ordered by position in hierarchy from high to low percentage endorsement.

4. Discussion

In the present study we investigated whether the ID screening tool for children, the CAIDS-Q, conformed to the properties of the MHM and DMM. We found that all of the CAIDS-Q item scalability coefficients were greater than the recommended minimum of .30. The total scale scalability coefficient was .60, which, according to Mokken's (1971) criteria, is indicative of a strong scale. No significant violations of manifest monotonicity were identified, suggesting that the assumption of latent monotonicity held for the scale in the present data.

By contrast, the assumption of the DMM of non-intersection of item response functions was violated, necessitating the rejection of the DMM for the present data. As a result, while the CAIDS-Q items exhibit the property of stochastic ordering by level of the latent trait, the ordering of the items is not invariant across individuals.

In practice, these results mean that the CAIDS-Q total scores can be used by clinicians or researchers to obtain a quick estimate of an individual's level of functional ability, without the need for a full and lengthy assessment. However, individual items cannot be used to approximate functional ability levels. In a clinical context the ability to use the scale total scores to estimate functional ability is likely to be particularly useful in situations where there may be a lengthy wait for formal diagnosis, but an early indicator of the potential support needs of a child is needed to allow planning for support to begin. It may be equally useful in forensic settings, where a quick and easy indicator of the functional abilities of a young person may allow potentially vulnerable young offenders to be identified and supported at an early stage (Ford et al., 2008). Similarly, researchers may find it useful to have an indication of functional ability in order to be able to match intellectually impaired individuals with controls, without the need for full diagnostic assessments that can be time-consuming (Winters, Collett, & Myers, 2005) and expensive to administer.

Conformity to the MHM also makes the tool potentially useful for tracking the developmental progress of an individual over time, if it is assumed that the between-individuals' observation of increasing scores indicating improvements in functional abilities also holds within individuals. Our sample included individuals with ID but also individuals with impairments that were not severe enough to merit this diagnosis. Our results are, therefore, not limited to individuals with ID, but apply to those children and young people who have below average levels of cognitive and adaptive functioning. Nevertheless, the possibility that scores increase over time for individuals as they develop, as they do across individuals will require further investigation, particularly given that individuals with more severe intellectual impairments may plateau at a relatively early developmental stage.

Although scale scores for the CAIDS-Q may, therefore, have utility beyond indicating the presence or absence of ID, the fact that the DMM did not fit in the present sample implies that individual items of the CAIDS-Q used in isolation are not necessarily a reliable indication of an individual's level of functional ability. There are several possible reasons why the DMM did not hold in the present study. First, it may be related to the selection of items that are too similar in terms of their difficulty or location along the continuum of the latent trait. Although this may be seen as a weakness in the current context, it may actually be a strength in the context of screening instruments, where it is desirable to maximise the amount of information the test conveys about trait levels in and around the level representing a diagnostic threshold. Thus, it is not surprising that the CAIDS-Q, which was designed as a screening tool for ID, would show a restricted range of item difficulties. This could result in item response functions being close together and making violations of their non-intersection more likely to happen because it is easier for an individual's pattern of responses to differ from that expected just due to chance. Indeed, in the present study, the percentage endorsement of items was generally quite similar, suggesting that this may have been the case. The current findings may be useful for future psychometric extensions of the CAIDS-Q, and suggest that items should be added that show a greater range of difficulty.

In a previous study, it was found that the adult version of the CAIDS-Q, the LDSQ, did in fact conform to the properties of the DMM, with the exception of 1 item (Murray & McKenzie, 2013). The 6 LDSQ items that conformed to the DMM showed greater variability in item endorsement than that observed in the present study and were endorsed 25%, 28%, 29%, 37%, 41% and 65% of the time respectively. The differences in variability in item endorsement may explain the discrepancy between

the DMM results in this and the LDSQ study. They suggest that the LDSQ may contain items that span a greater breadth of the functional ability continuum.

Another reason why the DMM may not have held in the present data is that developmental trajectories for different skills and abilities vary with ID aetiology. The CAIDS-Q includes items that are heterogeneous in content and could be argued to include skills in the social, visuo-spatial, perceptual, verbal and general intellectual ability domains. These various domains are known to be affected differently in various disorders that are related to ID. For example, social skills are especially impaired in those with ID when it is co-morbid with an autism spectrum disorder [ASD] (Matson, Mayville, Lott, Bielecki, & Logan, 2003), and it has been estimated by more recent research that between approximately 17–40% of individuals with ID also have ASD (see Matson and Shoemaker (2009), for an overview). This could mean, for example, that even if making friends was in general considered to be an easier task than telling the time, a person with ASD may experience the latter to be the easier task. This could manifest in Mokken scaling as violations of non-intersection for these two items in a number of individuals with a diagnosis of ASD. There is also evidence that the developmental trajectories of skills in different domains vary according to ID aetiology, providing further opportunities for such violations of non-intersection (Anderson, 1998). Note that although the CAIDS-Q has an item referring to having friendships, the adult version of the screening tool (the LDSQ), does not, making it less vulnerable to conflating severity of ASD with severity of intellectual impairment.

The DMM may also not have held because of the heterogeneity of the sample. It is expected that a person's level of functional ability would be both a function of age and of their IQ. This means additional scope for violations of the DMM, because, if the items do not form a consistent hierarchy, either across different age groups or across different IQ levels (or of ID severity), then it will not hold in a sample that is heterogeneous with respect to both.

In the present sample, we did not have sufficient data to conduct analyses separately for different aetiologies of ID, or to stratify by age or FSIQ, but this would be an interesting future study. Indeed, Mokken scaling is likely to be useful in determining the consistency with which skills in different domains are acquired in both children with neurodevelopmental disorders and in typically developing children.

Although recruitment difficulties and the relative infrequency of individuals with ID often restrict sample sizes, it would be a valuable future direction to collect sufficient data to conduct age and aetiology stratified analyses. Our results suggest that the CAIDS-Q can be used to estimate functional ability more generally in age heterogeneous samples. Collecting normative data for functional ability scores as measured by the CAIDS-Q at different ages would allow the scores for an individual to be compared with the expected score for an individual of their age, providing an estimate not only of their functional ability, but also how this translates into severity of ID. Such information would be highly informative, as functional ability in childhood would be expected to be strongly dependent on both age and the severity of the ID.

More generally, we believe that increased application of IRT based analyses can be useful in advancing the study of ID and below average functional ability in children. In the present study, we employed the technique to evaluate the scaling properties of the CAIDS-Q total scores and items, but the possibilities offered by IRT are much broader. It can also be used to investigate differential item functioning, in computer adaptive testing, and in placing items from different measures on to the same scale (Reise & Waller, 2009). For example, Wuang and Su (2009) used IRT based analyses to assess differential item functioning of a measure of visuo-motor integration across males and females with ID. Overall, however, the use of IRT in ID research has been fairly limited, in spite of a general trend for its increased use elsewhere.

Finally, it is important to note the limitations of our study. As we have noted above, our sample was not large enough to conduct stratified analyses. It would be useful in future studies to examine the scale stratified by age, sex, and aetiology in particular. In addition, the generalisability of our results is, thus far, limited to the kinds of participants included in the sample, that is, clinically referred children with low IQ on average.

In addition, although we have highlighted the benefits of the Mokken scaling procedure it also has some limitations. For example, parametric IRT models, though less flexible and requiring larger sample sizes, have the advantage over Mokken models of providing richer information on item performance.

5. Conclusion

We showed that the CAIDS-Q conforms to the properties of the MHM but not the DMM. This means that the scale can be validly used to estimate the level of functional ability of a child, rather than just providing a dichotomous 'yes', 'no' answer to the question of whether or not a child is likely to have ID.

References

American Academy of Pediatrics, Committee on Children with Disabilities. (2011). The pediatrician's role in the diagnosis and management of autistic spectrum disorder in children. *Pediatrics, 107*(5), 1221–1226.

Anderson, M. (1998). Mental retardation general intelligence and modularity. *Learning and Individual Differences, 10*(3), 159–178.

British Psychological Society. (2000). *Learning disability: Definitions & contexts.* Leicester: British Psychological Society.

British Psychological Society. (2003). *Issues in screening.* Leicester: British Psychological Society.

Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., et al. (2007). Efficacy of three screening instruments in the identification of autism spectrum disorder. *British Journal of Psychiatry, 191*, 554–559.

Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum Publishers.

Ford, G., Andrews, R., Booth, A., Dibdin, J., Hardingham, S., & Kelly, T. P. (2008). Screening for learning disability in an adolescent forensic population. *Journal of Forensic Psychiatry & Psychology, 19*, 371–381.

Goodman, J. F., & Linn, M. I. (2003). 'Maladaptive' behaviours in the young child with intellectual disabilities: A reconsideration. *International Journal of Disability, Development and Education, 50*(2), 137–148.

Guralnick, M. J. (2005). Early intervention for children with intellectual disabilities: Current knowledge and future prospects. *Journal of Applied Research in Intellectual Disabilities, 18*, 313–324.

Hamilton, S. (2006). Screening for developmental delay: Reliable, easy-to-use tools. *The Journal of Family Practice, 55*, 415–422.

Kingston, A., Collerton, J., Davies, K., Bond, J., Robinson, L., & Jagger, C. (2012). Losing the ability in activities of daily living in the oldest old: A hierarchic disability scale from the Newcastle 85+ study. *PLoS ONE, 7*(2), e31665.

McKenzie, K., & Megson, P. (2012). Screening for intellectual disability in children: A review of the literature. *Journal of Applied Research in Intellectual Disabilities, 25*(1), 80–87. http://dx.doi.org/10.1111/j.1468-3148.2011.00650.x

McKenzie, K., & Murray, A. L. (In press). The convergent validity of the Child and Adolescent Intellectual Disability Screening Questionnaire with a measure of adaptive functioning: A brief report. *Journal of Intellectual and Developmental Disabilities.*

McKenzie, K., Murray, A. L., Murray, K. R., & Murray, G. C. (2013). Assessing the accuracy of the WISC-IV seven subtest short form and the Child and Adolescent Intellectual Disability Screening Questionnaire in identifying intellectual disability in children. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence* http://dx.doi.org/10.1080/09297049.2013.799642 Advance online publication

McKenzie, K., Murray, G. C., & Murray, A. L. (2013). The validity of the Child and Adolescent Intellectual Disability Screening Questionnaire (CAIDS-Q) with children aged 6–7 years, 11 months: A brief report. *Psychiatry Research* http://dx.doi.org/10.1016/j.psychres.2013.06.014 Advance online publication

McKenzie, K., & Paxton, D. (2006). Promoting access to services: The development of a new screening tool. *Learning Disability Practice, 9*(6), 17–21.

McKenzie, K., Paxton, D., Murray, G., Milanesi, P., & Murray, A. L. (2012). The evaluation of a screening tool for children with an intellectual disability: The Child and Adolescent Intellectual Disability Screening Questionnaire. *Research in Developmental Disabilities, 33*(4), 1068–1075.

McKenzie, K., Paxton, D., Michie, A., Murray, G. C., Murray, A. L., & Curtis, J. (2012). Screening with young offenders with an intellectual disability. *Journal of Forensic Psychiatry and Psychology, 23*(5), 676–688. http://dx.doi.org/10.1080/14789949.2012.733723

Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities, 30*(6), 1107–1114.

Matson, J. L., Mayville, E. A., Lott, J. D., Bielecki, J., & Logan, R. (2003). A comparison of social and adaptive functioning in persons with psychosis, autism, and severe or profound mental retardation. *Journal of Developmental and Physical Disabilities, 15*(1), 57–65.

Maulik, P. K., & Darmstadt, G. L. (2007). Childhood disability in low- and middle income countries: Overview of screening, prevention, services, legislation, and epidemiology. *Pediatrics, 120*(Suppl. 1), S1–S55.

Meijer, R. R., & Baneke, J. J. (2004). Analyzing psychopathology items: A case for nonparametric item response theory modeling. *Psychological Methods, 9*(3), 354.

Mokken, R. J. (1971). *A theory and procedure of scale analysis.* The Hague: Mouton.

Murray, A. L., & McKenzie, K. (2013). Estimating the severity of intellectual disability in adults: A Mokken scaling analysis of the Learning Disability Screening Questionnaire http://dx.doi.org/10.1037/a0032369 Advance online publication

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5*, 27–48.

Ryan, J. J., Glass, L. A., & Brown, C. N. (2007). Administration time estimates for Wechsler Intelligence Scale for Children-IV subtests, composites, and short forms. *Journal of Clinical Psychology, 63*, 309–318.

Sijtsma, K., Meijer, R. R., & Van der Ark, A. L. (2011). Mokken scale analysis as time goes by: An update for scaling practitioners. *Personality and Individual Differences, 50*(1), 31–37.

Simonoff, E., Pickles, A., Chadwick, O., Gringras, P., Wood, N., Higgins, S., et al. (2006). The Croydon assessment of learning study: Prevalence and educational identification of mild mental retardation. *Journal of Child Psychology and Psychiatry, 47*, 828–839.

Sonnander, K. (2000). Early identification of children with developmental disabilities. *Acta Paediatrica Supplement, 434*, 17–23.

Stochl, J., Jones, P. B., & Croudace, T. J. (2012). Mokken scale analysis of mental health and well-being questionnaire item responses: A non-parametric IRT method in empirical research for applied health researchers. *BMC Medical Research Methodology, 12*, 74. http://dx.doi.org/10.1186/1471-2288-12-74

Van Der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software, 20*(11), 1–19.

Van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software, 48*(5), 1–27.

Watson, R., Van der Ark, L. A., Lin, L. C., Fieo, R., Deary, I. J., & Meijer, R. R. (2012). Item response theory: How Mokken scaling can be used in clinical practice. *Journal of Clinical Nursing, 21*, 2736–2746. http://dx.doi.org/10.1111/j.1365-2702.2011.03893.x

Wuang, Y. P., & Su, C. Y. (2009). Rasch analysis of the Developmental Test of Visual-Motor Integration in children with intellectual disabilities. *Research in Developmental Disabilities, 30*(5), 1044–1053. http://dx.doi.org/10.1016/j.ridd.2009.02.007

Winters, N. C., Collett, B. R., & Myers, K. M. (2005). Ten-year review of rating scales, VII: Scales assessing functional impairment. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*(4), 309–338.