Lawton IADL scale in dementia: can item response theory make it more informative?

Sarah McGrory1, Susan D. Shenkin2,3, Elizabeth J. Austin4, John M. Starr1,2,3

1Alzheimer Scotland Dementia Research Centre, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
2Geriatric Medicine, University of Edinburgh, Edinburgh, UK
3Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
4Psychology, University of Edinburgh, Edinburgh, UK

Address correspondence to: Sarah McGrory. Tel: 0131 650 4340; Fax: 0131 650 3461. E-mail: S.McGrory@sms.ed.ac.uk

Abstract

Background: impairment of functional abilities represents a crucial component of dementia diagnosis. Current functional measures rely on the traditional aggregate method of summing raw scores. While this summary score provides a quick representation of a person’s ability, it disregards useful information on the item level.

Objective: to use item response theory (IRT) methods to increase the interpretive power of the Lawton Instrumental Activities of Daily Living (IADL) scale by establishing a hierarchy of item ‘difficulty’ and ‘discrimination’.

Methods: this cross-sectional study applied IRT methods to the analysis of IADL outcomes. Participants were 202 members of the Scottish Dementia Research Interest Register (mean age = 76.39, range = 56–93, SD = 7.89 years) with complete itemised data available.

Results: a Mokken scale with good reliability (Molenaar Sijtsama statistic 0.79) was obtained, satisfying the IRT assumption that the items comprise a single unidimensional scale. The eight items in the scale could be placed on a hierarchy of ‘difficulty’ (H coefficient = 0.55), with ‘Shopping’ being the most ‘difficult’ item and ‘Telephone use’ being the least ‘difficult’ item. ‘Shopping’ was the most discriminatory item differentiating well between patients of different levels of ability.

Conclusions: IRT methods are capable of providing more information about functional impairment than a summed score.
‘Shopping’ and ‘Telephone use’ were identified as items that reveal key information about a patient’s level of ability, and could be useful screening questions for clinicians.

**Keywords:** item response theory, Mokken, dementia, functional measures, Lawton IADL, older people

---

**Introduction**

Functional impairment is a core feature of dementia. Functional impairment is widely measured using ‘Activities of Daily Living’ (ADL) scales which encompass the assessment of both Basic Activities of Daily Living and Instrumental Activities of Daily Living (IADL). The former include activities such as dressing, bathing and feeding, while the latter concern more complex activities such as handling finances, taking medication and doing housework. IADL items requiring more complex neuropsychological organisation are highly dependent on adequate cognitive capacity and are therefore most susceptible to the early effects of cognitive decline [1]. Assessing IADL can consequently be useful in detecting and diagnosing early dementia [2].

Many IADL scales have been developed and they are conventionally scored by summing the responses to individual items on the scale to yield a total score. Despite the popularity of this method, there are issues that make it difficult to interpret [3]. For example, the total-score method weights each item equally, which assumes that all items represent equal levels of severity. This is very rarely the case [4]: e.g. doing laundry is more challenging than feeding [5]. Furthermore, the total-score method asserts that each item on the scale is equally related to the construct under examination which again is rarely the case [4]. These limitations can be overcome using item response theory (IRT) methods [6, 7].

According to IRT, the items on a scale are related to a latent construct; functional impairment in the case of the current study. IRT is based on the probability of a person achieving a particular score on a test given their standing on the latent construct [8]. This better reflects the underlying trait than traditional methods [9]. A more in-depth discussion of IRT is beyond the scope of this article, for further information see Hambleton and Swaminathan [6].

IRT provides two useful measures; item ‘difficulty’ and ‘discrimination’. ‘Difficulty’ refers to the ability level necessary in terms of the latent construct for an individual to have a 0.5 probability of responding positively to a specific item. In the context of IRT, an item is considered ‘difficult’ if a high degree of ability is required in order to respond positively. Only those with a high level of ability will be able to endorse the ‘difficult’ items, whereas most will endorse or respond positively to the less ‘difficult’ items. From a clinical point of view ‘difficulty’ can be thought of as severity, for example the degree of functional impairment required to cause challenges with handling finances. With regards to IADL items the more ‘difficult’ a task is, the better the person’s functional ability must be in order to be able to perform the task. A hierarchy of item ‘difficulty’ details the expected order of functional impairment. Hierarchies of ADL/IADL scales have been confirmed using IRT methods [5, 10, 11]. These hierarchies found that items differed in terms of ‘difficulty’ with items such as ‘Shopping’ and ‘Doing Laundry’ being more ‘difficult’ than ‘Dressing’ and ‘Transferring’ [5] and ‘Active recreation’ and ‘Volunteer job’ being more ‘difficult’ than ‘Taking care of health’ and ‘Personal care needs’ [10]. However, these studies did not examine functional impairment within dementia populations. Hierarchies outlining the functional impairment in dementia would offer prognostic value to researchers and clinicians investigating functional impairment by identifying any deviations in the rate of decline from the typical trajectory of loss [11].

‘Discrimination’ is the extent to which the item distinguishes participants with relatively low functional ability from those with relatively high levels of ability. An item with poor ‘discrimination’ will distinguish poorly between mild and severe levels of functional impairment because the probability that the person will endorse the item is nearly the same across all levels of severity. An item with good ‘discrimination’ distinguishes well between varying levels of functional ability because as the level of severity increases so too does the probability that a respondent will be unable to perform the task. For example, Fico et al. [11] determined ‘Prepare a meal’ had very weak discriminatory value and did not differentiate between people of different abilities, whereas ‘Get on a bus’ was the most discriminatory item differentiating between those with low functional ability and those with high functional ability. Determining item ‘discrimination’ can identify key items on a scale and highlight weaker items or those whose function is redundant [12].

IRT methods have been applied to ADL/IADL scales in general populations [5, 10, 11]. However, the Lawton IADL scale has not been analysed with IRT methods to investigate the pattern of functional impairment caused by dementia. This analysis in a sample comprises people with dementia could provide clinically useful information. Therefore, this study applied IRT methods to the Lawton IADL scale to establish a hierarchy of item ‘difficulties’ and to assess the ‘discriminatory’ power of the items in people with dementia.

**Methods**

**Sample**

Data were obtained from the Scottish Dementia Research Interest Register, described in detail previously [13]. Participants
were referred by a clinician, and had a diagnosis of dementia or related cognitive disorder. Two hundred and two participants for whom full itemised IADL data were available were included in the current study. Participants were assessed by clinical studies officers trained to ensure consistency.

**Measures**

The Lawton IADL scale [14] is a self-report scale that assesses eight tasks providing information about functional skills necessary to live independently in the community, i.e. the ability to use the telephone, shop, prepare food, handle finances, do housework, take medications, do laundry and travel. Each activity can be scored as either 1 (can perform task independently) or 0 (not able to do).

**Statistical analysis**

Non-parametric item response theory was used to confirm a hierarchy of item ‘difficulty’ for the Lawton IADL Scale, and to establish the ‘discriminatory’ power of each item in the scale. Data were analysed using the Mokken scaling analysis (MSA) package in the public domain software ‘R’ [15]. Mokken is a non-parametric method based on IRT principles for determining whether the items form a hierarchy based on mean scores.

There are two probabilistic models in Mokken scaling; the monotone homogeneity model (MHM) and the double monotonicity model (DMM). Both models relax some of the strict assumptions of traditional parametric models such as Rasch or two-parameter logistic model. These models inspect item scores relative to one another to see how consistently they are ordered.

The MHM assumes unidimensionality, local independence and latent monotonicity of the items within a scale. When these assumptions are met, people can be ordered with respect to their latent trait value by the unweighted sum of their total item scores. The MHM seeks unidimensional sets of items based on scalability coefficients developed by Loevinger [16]. Coefficients for each item in the scale (Hi) are calculated to determine whether the item is coherent enough for inclusion in the scale. For a unidimensional scale, all Hi’s must be >0.3 [17]. Once the fit of the MHM to the data has been determined Hi can be interpreted as an index of item ‘discrimination’ with higher values indicating greater ‘discrimination’ [18].

Hi coefficients are used to calculate an overall scale coefficient (Loevinger’s H). Loevinger’s H is a measure of the strength and quality of the Mokken scale and reflects the extent to which the items appear in the same relative order [19]. Higher values of H imply fewer violations and a stronger hierarchy [20]. A violation is defined as any deviation from the expected ordering. Scales with H < 0.3 are not considered unidimensional, H < 0.4 denotes a weak scale, H < 0.5 represents a medium scale and H > 0.5 is a strong scale [16]. When the assumptions of the MHM met people can be ordered with respect to their latent trait value by the unweighted sum of their total item scores.

The DMM includes the assumptions of the MHM plus the assumption of non-intersection of items within a scale. If the assumptions of the DMM hold, this implies the items are invariantly ordered. Invariant item ordering (IIO) means the items have the same ordering of ‘difficulty’ irrespective of the value of latent trait. IIO is necessary for confirming hierarchical scales that are replicable across samples. The ‘difficulty’ of the items should remain in the same order for all respondents regardless of their latent trait value [21]. To investigate IIO, the ‘check.iio’ function of the MSA package in ‘R’ was used. IIO is established by Htrans (Htrans) which is the equivalent to H in terms of the same minimum values signifying weak, moderate and strong accuracy of item ordering in the scale [22]. The reliability of Mokken scales can be estimated using the Molenaar Sijtsma (MS) statistic [23]. MS gives a direct estimate of the reliability of a test score with MS > 0.7 taken to indicate a reliable scale [24].

**Results**

The sample comprises 202 participants (105 male), mean age 76.39 years (SD = 7.90, range = 56–93), mean Mini Mental State Examination (MMSE) [25] score 22.10 (range = 4–30, SD = 5.05, median = 23, interquartile range = 6). A variety of aetiologies were included: Alzheimer’s disease (AD) (133), mixed AD/vascular disease (VaD) (35), VaD (17), frontotemporal dementia (2), dementia with Lewy bodies (4), Parkinson’s disease dementia (4), mild cognitive impairment (6) and one with uncertain diagnosis.

A single scale with MS = 0.79 was obtained which indicates a reliable scale and satisfies the IRT assumption of a single unidimensional scale. The overall Loevinger’s H value of 0.55 indicates a strong scale. Of the eight items included in the Lawton IADL scale ‘Shopping’ was the most ‘difficult’ (mean score of 0.21) (see Supplementary data available at Age and Ageing online, Appendix 1, Table 1). ‘Telephone use’ was the least ‘difficult’ (mean score of 0.92). Supplementary data available at Age and Ageing online, Table S1 also show the Hi coefficients for each item.

Item scalability coefficients (Hi) were all positive and clearly exceeded the 0.3 threshold, signifying that the items meet MHM assumptions. Items can be ordered in terms of ‘discrimination’ from high to low by these Hi’s: ‘Shopping’ was the most ‘discriminatory’ and ‘Travel’ the least. The Lawton IADL scale showed IIO (Htrans0.64). This indicated that the ordering of the items is the same for all levels of the latent trait, i.e. the items are in the same order of ‘difficulty’ regardless of the severity of functional impairment.

**Discussion**

These data provide significant, novel information about the validity and practical worth of the Lawton IADL scale in this sample of people with dementia. It is a strong unidimensional functional ability scale, with H of 0.55 and MS reliability statistic of 0.79. ‘Shopping’ and ‘Food preparation’ were found to be the most ‘difficult’ items and therefore those lost earliest in the disease process. ‘Telephone use’ was the
least ‘difficult’ item with problems performing this task indicating severe impairment. A patient reporting challenges with ‘Telephone use’ is very unlikely to be able to perform any other task in the scale. Likewise, it is likely that a patient reporting no problems with ‘Shopping’ or ‘Food preparation’ will have no limitations with other tasks.

These findings have useful clinical implications. People requiring assistance with the most ‘difficult’ item ‘Shopping’ should alert clinicians as, in the context of cognitive decline, it could herald the initial phase of functional impairment. Problems performing complex activities of daily living have been reported to precede dementia diagnosis by as much as 10 years [26]. As the items of the Lawton IADL scale conform to a formal hierarchy the most ‘difficult’ items such as ‘Shopping’ and ‘Food preparation’ can act as sensitive indicators of impending disability in the other activities [27].

Items with high ‘discrimination’ are better able to detect differences in effects of interventions or drug therapies [28]. Ideally, a measure should comprise items of differing degrees of ‘difficulty’ right across the spectrum of ability and demonstrate high levels of ‘discrimination’. This ensures that changes at every point along the ability spectrum will be detected resulting in more reliable and accurate measurement.

The inclusion of items such as ‘Shopping’ and ‘Food preparation’ which showed high ‘discrimination’ may assist in the detection of small changes in milder stages of dementia as these abilities are lost rapidly at an early stage. ‘Telephone use’ discriminates well at the lower end of the hierarchy. The creation of more items such as this may help to introduce greater ‘discrimination’ in the more advanced stages. IRT analyses can be applied to IADL/ADL scales making them more sensitive to identifying and monitoring changes in both mildly and severely impaired patients. Better assessment of the rate of decline could enhance prediction of future deterioration.

The study was predominantly restricted to patients with mild–moderate dementia (mean MMSE score 22.1, SD = 5.05) with a range of aetiologies. Future research should investigate the loss-of-functional independence in more severe samples, and in specific dementia subtypes. For example, there is more rapid deterioration of functional abilities in patients with Frontotemporal dementia compared with Alzheimer disease [29]. The majority of this sample (80%) was taking Cholinesterase Inhibitors which are acknowledged to be effective in delaying or slowing the worsening of symptoms, although these effects are not large [30].

IRT has benefits not only in the monitoring of patients, but establishing the sequence of decline which can also help in characterising adaptations to disability and differences between subgroups. While additive summary scores can be helpful in summarising overall function, they can conceal as much information as they reveal, and IRT methods are a useful method to increase the information provided by simple functional scales.

Furthermore, simultaneous analyses of cognitive and functional scales could enable the discovery of more precise associations between cognitive and functional outcomes.

Key points

- IRT analyses provide more information than the summed score method.
- The ability to shop and to use the telephone was identified as key items which could be valuable screening questions for clinic.
- IRT analyses can make IADL scales more sensitive to identifying and observing changes in functional abilities in dementia.

Conflicts of interest

None declared.

Funding

This work was supported by a PhD studentship from Alzheimer Scotland.

Supplementary data

Supplementary data mentioned in the text are available to subscribers in *Age and Ageing* online.

References

7. Fieo RA, Austin EJ, Starr JM, Deary IJ. Calibrating ADL-IADL scales to improve measurement accuracy and to extend the disability construct into the preclinical range: a systematic review. BMC Geriatr 2011; 11: 42–57.
9. Chan KS, Kasper JD, Brandt J, Pezzin LE. Measurement equivalence in ADL and IADL difficulty across international


16. Loevinger J. A systematic approach to the construction and evaluation of tests of ability. Psychol Monogr 1947; 61, No.4.


27. Finlayson M, Mallinson T, Barbosa VM. Activities of daily living (ADL) and instrumental activities of daily living (IADL) items were stable over time in a longitudinal study on aging. J Clin Epidemiol 2005; 58: 338–49.


Received 21 December 2012; accepted in revised form 4 October 2013