Journals of Gerontology: MEDICAL SCIENCES
Cite journal as: J Gerontol A Biol Sci Med Sci 2014 June;69(6):766–772
doi:10.1093/gerona/glt167

Advance Access publication October 22, 2013

Functional Status in the Young–Old: Establishing a Working Prototype of an Extended-Instrumental Activities of Daily Living Scale

Robert Fieo,[1] Jennifer J. Manly,[1] Nicole Schupf,[2] and Yaakov Stern[1]

[1]Cognitive Neuroscience Division, Department of Neurology and Taub Institute and
[2]The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University College of Physicians and Surgeons, New York.

Address correspondence to Prof. Yaakov Stern, PhD, Cognitive Neuroscience Division, Department of Neurology and Taub Institute, Columbia University College of Physicians and Surgeons, New York, NY 10032. Email: ys11@columbia.edu

Background. Instrumental activities of daily living (IADLs) exhibit strong predictive power for the presence of dementia and mild cognitive impairment. However, IADLs are often less effective in younger cohorts or in healthy community-dwelling samples, presenting with large ceiling effects. This study aimed to construct an IADL scale with an extended range. An effort was made to incorporate leisure activity tasks that were more stimulating, and potentially more challenging, into a set of traditional IADLs.

Methods. Beginning with a set of IADL and leisure activity items, nonparametric item response theory methodology was used to construct a scale with appropriate dimensionality, monotonicity, item discrimination power, and scalability within a large cohort of young–old (aged 65–75). Dimensionality was further scrutinized by principal component analysis of the residuals. The predictive validity of the resulting scale for poor cognitive performance was evaluated using logistic regression.

Results. A reliable (ρ = .73) unidimensional construct was established, meeting the Mokken item response theory criteria of medium scalability. Excluding demented participants, the adjusted model proved sensitive to relatively subtle cognitive deficits; each additional task endorsed (nine-item scale) significantly decreased the odds of being in the bottom quarter of composite domains relating to processing speed (odds ratio = 0.73 [confidence interval: 0.56–0.97], $p < .05$) and visuospatial ability (odds ratio = 0.70 [confidence interval: 0.73–0.87], $p < .01$).

Conclusions. A reliable extended-IADL scale was constructed meeting item response theory assumptions relating to unidimensionality, monotonicity, and invariant item ordering. The range of measurement extends well beyond traditional IADL scales. Finally, the scale appears to be sensitive to cognitive differences within the normal spectrum.

Key Words: Cognitive aging—Functional performance—Successful aging.

Received May 15, 2013; Accepted September 2, 2013

Decision Editor: Stephen Kritchevsky, PhD

Introduction

Decline in instrumental activities of daily living (IADL; ref. 1) is often a gradual and progressive deterioration across a number of tasks many years before the onset of dementia (2), with IADL impairment being described as a "stepping stone towards Alzheimer's disease diagnosis in participants with mild cognitive impairment" (3,4). However, incorporating IADLs into a multifaceted approach to early detection of cognitive changes in healthy aging is limited by the fact that IADL outcomes present with large ceiling effects when used in community-dwelling populations. For example, Wang and colleagues (5) reported disability rates for community-dwelling men and women (mean age, 72) to be 17% in the IADL domain. In a longitudinal population-based cohort (n = 4,287 initially nondisabled older adults from a biracial urban community) with a mean follow-up > 9 years, the average age of IADL disability was 79 (6). Furthermore, it is evident that IADL disability rates are driven primarily by the physical IADLs, as differentiated by Ng and colleagues (7), rather than cognitive IADLs such as using the telephone, personal finance, and medication management. For instance, in a community sample with a mean age of 80 endorsing difficulty in heavy housework was 30%, but difficulty in cognitive IADLs was endorsed at a rate of 2% for paying bills and 3% for telephone use (8). These values fall well short of standards relating to content validity; a maximum ceiling effect of 15% for any given sample has been proposed as the reasonable limit, with some investigators proposing as low as 10% (9).

In the same way that IADLs have been operationalized as exhibiting a greater level of complexity than basic ADL (10), there should be a set of activities that are potentially more challenging or stimulating than traditional IADLs, for example, attending classes or playing an instrument. Here, we considered participation in various social and intellectual leisure activities. In cross-sectional and prospective longitudinal studies, reduced engagement in these activities is associated with increased risk of prevalent or incident Alzheimer's disease (11,12).

The first aim of this study was is to ascertain whether performance in leisure activities can be placed on a continuum along with traditional IADL tasks. The second aim was to determine whether the summed score of a new scale comprised of these items is an accurate reflection of changes in the latent construct relating to cognitive function. The final aim was to assess the scales association with cognitive ability, using linear and logistic regression methods.

Methods

Participants

The study participants were collected from a probability sample of Medicare beneficiaries living within three adjoining census tracks in the northern Manhattan (New York City) communities of Washington/Hamilton Heights and Inwood. The sample included individuals from several countries of origin and three broadly defined ethnic categories—Caribbean Hispanic, black, and non-Hispanic white of European ancestry. Those elders not speaking English or Spanish were excluded from participation. The study included longitudinal data from two separate recruitment periods, one that began in the 1992 and the other in 1999. The sampling strategies and recruitment outcomes of these two cohorts are described in previous publications (13). Reassessments took place in waves that were approximately 18–30 months apart.

A central aim of this study was to develop a scale that was sensitive to functional status (ie, leisure activities grounded in IADLs) in the young–old, thus the sample was restricted to participants aged 65–75 (n = 1,430), with a mean age of 71, *SD* 2.5. A majority of the sample was seen only once when within this age range, with 41% of participants evaluated at least twice, and 12% of the sample evaluated ≥3 times. For scale development, multiple assessments from the same individual were not excluded, thus the overall scale analysis incorporates intraindividual variability within the sample. All participants presenting with dementia from ages 65–75 were removed form analysis, as the scale was intended for use in relatively healthy, or preclinical participants. During scale analysis or construction, 8% of the sample presented with mild cognitive impairment amnestic and another 10% were mild cognitive impairment nonamnestic. After scale construction, was examined the relationship between the extended-IADL (E-IADL) scale and cognitive ability in a smaller sample of participants, all aged 70 (n = 214). Demographics for both samples are presented in Table 1. We chose to limit our sample to 70-year-old participants to highlight the challenge of detecting cognitive deficits in younger participants, as the robust correlation of age and cognition is often observed in gerontology literature. Further, we sought to draw attention to the limitations of using traditional IADL items in younger community-dwelling older adults.

Table 1. Demographics

	Sample Characteristics During Scale Construction (n = 1,430)	Sample Characteristics of Regression Analyses (n = 214)
Mean age	71 (*SD* 2.4)	70
Women %	66	63
Race		
Hispanic	42	40
Black	31	34
White	26	24
Mean education	10.3 (*SD* 4.7)	10.6 (*SD* 4.5)
Occupation		
High (%)	25	26
Low (%)	71	72
Housewife (%)	9	3
Depression (%)	21	19
Repeated measures		
Percentage with two evaluations	28	—
Percentage with three to four evaluations	14	—

Notes: Percentage depressed reflects scores ≥4 on the Center for Epidemiologic Studies–Depression scale.

High occupation = manager business/government and professional/technical. Low occupation = unskilled/semiskilled, skilled trade or craft, and clerical/office worker.

Statistical Analyses

Construct validity.—We used the public domain software "R" for Mokken analysis (14) to examine the scalability of leisure activities presented in Table 2. The Mokken model of monotone homogeneity (MH) is based on the assumptions of unidimensionality, local independence, and monotonicity in the latent attribute. Unidimensionality is achieved when all test items or tasks support a single underlying dimension. Local dependency occurs when a respondent's answer to one item automatically has a bearing upon the answer to the other. The local independence assumption seeks to confirm that there is no significant association between items after the dominant factor (eg, functional ability) has been conditioned out (15). Finally, monotonicity ensures that item endorsements increase as trait levels increase.

The MH model produces a "scalability" diagnostic (16) to assess these assumptions (17). The primary function of

Table 2. Original Leisure and IADL Sets With Response Options

Original Leisure-Type Items With Response Formats	Original IADL Items Used With Response Formats
In the last month, have you (done)…	1) Did you see the phone in the last month? If yes, do you dial the number by yourself, or does someone help you? Response: No help used (0); Used help (1); Didn't use the phone (2).
1) Any knitting, played any musical instrument, or spent time on any other hobby?	2) In the past month, how much of the cooking and preparation of meals did you do yourself?
2) Gone to visit friends or relatives?	3) Do you do all or almost all of your shopping on your own without help from anyone else? Response: All or most all by self (0); Most by self, others do rest (1); Some by self, others do the rest (2); None or almost nothing by self (participation with assistance or supervision); Not applicable (9).
3) Received visits from relatives or friends at your home?	4) Do you have any difficulty doing light chores such as washing dishes, cleaning the stove top and kitchen counter, and taking out the garbage by yourself? Response: No (0); Yes (1); Not applicable (9).
4) Gone out to a movie, restaurant, or sporting event?	5) Do you take medications without help from anyone? Or does someone remind you, help you, or give them to you? Response: Takes all medications by self (0); Someone reminds of when or how much (1); Someone gives medications (2); Not applicable (9).
5) Read magazines, newspapers, or books?	6) How much trouble does the participant have doing chores around the house, like cleaning, doing the laundry, cooking, mowing the lawn, or other chores?
6) Watched television or listened to radio?	7) How much trouble does the participant have handling money, like counting change at a restaurant, paying cab fares or paying for groceries?
7) Unpaid community/volunteer work?	8) How much trouble does the participant have finding their way around the house?
8) Played cards/games/bingo?	9) How much trouble does the participant have finding their way around the neighborhood? Response: None (0); Some (1); A lot (2).
9) Gone to a club or center to take part in activities?	
10) Gone to classes of any kind?	
11) Gone to church/synagogue/temple services?	
Response options: No (2); Yes (1); Not applicable (9)	

Notes: All IADL response options were collapsed to a binary response option so as to make them compatible with dichotomous leisure activities, that is, some trouble or difficulty versus no help. IADL = instrumental activities of daily living.

the Mokken scalability component is to assess the degree to which participants can be accurately ordered on the latent trait by means of their sum score (18); MH allows for the ordering of persons on the latent trait by the sum of the item score—an essential requirement for a psychological test. Item quality is expressed by discrimination power. Poorly discriminating items have response probabilities that are less responsive to changes in trait level. From the item *H* values, the Mokken model produces a weighted sum, referred to the total *H* or scale *H*. When interpreting *H*, the following guidelines are common: 0.3–0.4 = weak scale, 0.4–0.5 = medium scale, and > 0.5 = strong scale (19). Unidimensionality and local independence were further assessed using the RUMM2030 software package (20). Here, unidimensionality and local independence were scrutinized by examining principal component analysis of the residuals (21) and the residual correlations for each pair of items, where values above .2 indicate local dependence (22).

Finally, for the nonparametric Mokken scaling, Rho is used to define scale reliability, and is an internal consistency coefficient comparable to Cronbach's alpha (23). Most theorists agree that a ρ more than .80 is desirable, and a ρ more than .70 is a minimum requirement (24).

Cognitive associations.—Data analysis was performed using SPSS version 16.0 (SPSS Inc., Chicago, IL). A Spearman's rank order correlation was run to determine the relationship between the E-IADL sum score and the four composite scores of memory, language/executive, speed, and visuospatial (see Supplementary Material). Correlations between composite scores and IADL sum scores were also examined. Frequency distributions of E-IADL and IADL sum scores were generated as evidence of content validity. Predictive validity was investigated cross sectionally for a sample of nondemented participants aged 70 (n = 216), using a binary logistic regression model. E-IADL sum was used as a continuous predictor and performance on neuropsychological composites as the outcome measure. The neuropsychological composites were dichotomized into the 25% worst performers versus all those above this cut point.

Results

Scale Construction

Stage 1.—The initial examination of 11 leisure activities revealed a set of items that were not scalable under the Mokken MH criteria. The overall scale *H* coefficient was 0.25, and nearly half of the item *H* coefficients fell below the lower band value of 0.30. Furthermore, two items, "receive visitors" and "watch TV," needed to be excluded from further analysis due to negative *H* values with one or more items. Three additional items ("games," "hobby," and "attending church") were removed due to lowerbound *H* values under the 0.30 standard. The six remaining items ("classes," "volunteer," "clubs," "movies," "visiting," and

"reading") did meet the Mokken MH criteria, resulting in an overall scale *H* coefficient of 0.45 (see Table 3). This *H* value suggests that the scale has medium strength with regard to ordering participants by the sum of items and approaching the 0.50 cut point for a strong scale.

Stage 2.—In an effort to reduce the often dramatic ceiling effects observed for the young–old being assessed on IADLs and to anchor complex leisure activities in more fundamental tasks, an attempt was made to combine the two scales into a single unidimensional construct. This resulted in low scalability ($H = 0.39$) for 13 activities. The Mokken package flagged the "paying bills" item due to a lowerbound *H* value of 0.26. This item was removed from further analyses. In order to improve the overall scalability, the worst performing leisure item ("reading"; $H = 0.30$) and the worst performing IADL item ("trouble around the house"; $H = 0.32$) were removed. Low item *H* values infer weak discriminatory power for distinguishing between people with low functioning and those with high functioning. Removing these items improved the overall scalability to 0.41, indicating medium scale strength, suggesting that we can be reasonably confident that the sum of these items reflects changes at the latent trait level.

Stage 3.—Further investigations of unidimensionality and local independence of the remaining 10 items were performed through principal component analysis of the residuals. The first principal components analysis resulted in two-item subtests formed by the five most positive loading items (shopping: loading of 0.569; light chores: 0.545; prepare meals: 0.520; trouble in neighborhood: 0.279; medication use: 0.274) versus the five most negative loading items (classes: loading of –0.471; clubs: –0.453; volunteer: –0.391; movies: –0.229; visit friends: –0.097). Comparing the subtests through an independent *t* test demonstrated that only 3.99% of participant ability estimates differed significantly, which supports the unidimensionality of the E-IADL. The local dependency assumption was investigated by computing the residual correlations for each pair of items. Here, we found no correlations that were .2 above the average correlation for the entire matrix, thus indicating no violation of local independence.

Reliability was estimated using the Mokken Rho coefficient, an internal consistency coefficient comparable to Cronbach's Alpha (23). The value of 0.73 indicates sufficient reliability; in psychometrics (24) as well as in medical statistics (25) values of 0.70 or 0.80 are considered to be sufficient for basic research, such as comparing groups and calculating correlations.

Stage 4.—In addition to the MH scalability procedure, the Mokken model can be used to apply the Double Monotonicity procedure. Double monotonicity is used to establish the property of invariant item ordering (IIO), thus confirming a formal item hierarchy. Inspecting item characteristic curves revealed that several curves crossed, suggesting that the item hierarchy was inconsistent over the latent construct. The "prepare meals" item was involved in both violations and was removed to meet

Table 3. Original Items and Those Meeting Mokken Scalability Criteria

Original Item Set IADLs	Items Meeting Mokken Scalability Criteria IADLs	Mean	Item *H*
Difficulty with light chores	Difficulty with light chores	0.88	0.49
Shopping	Shopping	0.89	0.49
Preparing and cooking meals	Preparing and cooking meals	0.90	0.44
Assistance with paying bills	Assistance with paying bills	0.94	0.31
Trouble around the neighborhood	Trouble around the neighborhood	0.95	0.43
Help with medication	Help with medication	0.97	0.45
Trouble around the house	Trouble around the house	0.98	0.41
Trouble with chores	—		
Used help dialing telephone in month	—		
	Scale H: 0.41		
Cognitive Leisure Activities	**Cognitive Leisure Activities**		
Gone to classes of any kind	Gone to classes of any kind	0.14	0.48
Community volunteer work	Community volunteer work	0.24	0.41
Club or center activities	Club or center activities	0.27	0.44
Movies, restaurant, sporting event	Movies, restaurant, sporting event	0.57	0.48
Visiting friends or relatives	Visiting friends or relatives	0.79	0.42
Read magazines, newspapers, or books	Read magazines, newspapers or books	0.87	0.45
Play cards/games/bingo	—		
Knitting, musical instrument, or other hobby	—		
Church/synagogue/temple	—		
Receive visitors in home	—		
Watch TV or listen to radio	—		
	Scale H: 0.45		

Notes: *H* coefficients within the scale represent item discriminatory power. IADL = instrumental activities of daily living.

Table 4. Final E-IADL Scale

E-IADL Scale	Mean	Item *H*	*SE*
Gone to classes of any kind	0.14	0.46	.025
Community volunteer work	0.24	0.40	.021
Club or center activities	0.27	0.42	.021
Movies, restaurant, sporting event	0.57	0.44	.020
Visiting friends or relatives	0.79	0.36	.022
Doing light chores	0.88	0.43	.023
Shopping	0.89	0.43	.023
Trouble around neighborhood	0.95	0.38	.034
Help with medication	0.97	0.45	.038
Scale H: 0.42, $H^T = 0.78$			

Notes: Scale and Item *H* Statistics reflect Loevinger's homogeneity coefficient. H^T = H coefficient transposed. IADL = instrumental activities of daily living.

the IIO criteria. The final nine-item E-IADL scale (Table 4) met the minimum reliability requirements and a scalability coefficient of 0.42, which indicates a scale of medium strength.

Cognitive Associations

One point was given for participation in each of the E-IADL (nine-item scale) activities and an aggregate score was assigned to each patient for the subsequent analysis. The correlation coefficient (Spearman's rho) for the E-IADL sum score and each cognitive domain were: visuospatial ($r = .35$, $p < .01$), processing speed ($r = .36$, $p < .01$), memory ($r = .14$, $p < .05$), language function ($r = .35$, $p < .01$). To further scrutinize the E-IADL, we contrasted its performance with a nine-item sum score derived from traditional IADLs. The IADL items included the initial items presented in Table 3: difficulty using the telephone, taking medications, light chores, paying bills, trouble around the house, trouble around the neighborhood, difficulty preparing and cooking meals, and difficulty shopping, and trouble with chores. Comparing the association between the four cognitive composite scores and IADL sum score resulted in one significant correlation (rho), visuospatial ($r = .18$, $p < .05$). Directly comparing the IADL to E-IADL scores in terms of linear regression suggest that the E-IADL accounts for a much larger proportion of variance in visuospatial performance –0.09 versus 0.03.

Multivariate binary logistic regression was employed to examine whether the E-IADL sum score could be used to differentiate between participants who fell in the bottom 25% of cognitive performance versus all of those above this mark. This dichotomization was established for three of the four composite measures: visuospatial ability, speed of processing, and language function. Here, we found that one unit increase in the scale significantly decreased the odds of being in the bottom 25% of visuospatial performance (odds ratio = 0.70 [confidence interval: 0.54–0.89], $p < .01$) and processing speed (odds ratio = 0.71 [confidence interval: 0.54–0.93], $p < .05$). Covariates in the model included education, occupational status, gender, ethnicity, and depression. Despite having a significant univariate association with the E-IADL (odds ratio = 0.78 [confidence interval: 0.62–0.98], $p < .05$), the language composite was no longer significant after the inclusion of the education variable. However, dividing the E-IADL into high and low performers did result in a significant association. Because the E-IADL scale is composed of four traditional IADL items and five leisure-type activities, higher performers were designated as those who could perform from 1 to 5 leisure activities and low performers, those participants only performing from 1 to 4 of the traditional IADL items. Dichotomizing activities in this way resulted in an odds ratio 2.96 (confidence interval: 1.23–7.14, $p < .05$), that is, low performers being in the bottom 25% of language scores.

Discussion

Prompted by historical evidence indicating that more complex IADLs can be placed on a unidimensional continuum with basic or personal care ADL (26,27), this study hypothesized that cognitive leisure activities might occupy the challenging end of a continuum anchored in less stimulating IADL tasks. Unidimensionality was confirmed for a nine-item scale using two distinct methods, the Mokken MH criteria and principal component analysis of the residuals. The final scale presented with medium scalability, as quantified with an *H* coefficient of 0.42. Binary logistic regression indicated that participants only performing IADLs, compared with those performing one or more leisure activities, were 2.96 times (relative odds) more likely to be in the bottom 25% of the language composite distribution.

Cognitive impairment without dementia has been of interest to clinicians because of the difficulties they may foster in performing tasks of everyday function. Compared with basic ADLs (eg, bathing or dressing), IADLs such as shopping or preparing meals generally have greater cognitive demands and are in consequence more likely to be vulnerable to early effects of cognitive decline (28). In contrast, traditional IADLs may be less effective at early phases of cognitive impairment, allowing transitions from healthy cognitive aging to neurodegenerative diseases to go unnoticed during critical intervention periods. The E-IADL is intended to act as a more demanding functional marker, or functional risk factor, compared with IADLs.

The E-IADL sum score correlated most strongly with processing speed, visuospatial ability, and language function. In regression analyses, processing speed and visual spatial ability exhibited the most robust relationships with E-IADL performance. Newson and Kemps (29), using a large general population sample (mean age, 77), examined volunteer work and social activities and found that these tasks significantly predicted 6-year cognitive change in processing speed. In a population-based sample (mean age, 69), Aartsen and colleagues (30), found that processing speed was the only cognitive score that influenced engagement in their domain of development activities, as measured by two activities—following an educational course and engaging

in outdoor sports. In the current study, using cross-sectional data of participants aged 70, it was found that for each additional E-IADL task endorsed the odds of being in the bottom quarter of a processing speed composite decreased by 30%. A nearly identical association was found for the domain of visuospatial ability. In a community-dwelling sample with a mean age of 55 (31), it was found that a spatial visualization composite was significantly associated ($r = .29$, $p < .01$) with their developmental activities domain (eg, participate in college course, read magazine for job, attend public lectures).

This manuscript is largely a methodological endeavor so we would like to make two final points concerning scale construction and item response theory (IRT) standards. Local independence has been proposed as one of several criteria in establishing construct validity (32). Most often researchers assume local independence, without inspection. Unchecked local dependence can affect parameter estimates, standard error estimates, and the fit of the IRT model (33). Our final point concerns the property of IIO. Ligtvoet and colleagues (34) report that IIO is a strong requirement in measurement practice, and that researchers sometimes assume that fitting an IRT model implies that items have the same ordering by difficulty or popularity for all individuals, but this assumption requires modification. Furthermore, previous research has shown (35,36), rather surprisingly, that IIO can only be confirmed in a small minority of IRT models.

Study Limitations

It is possible that participants do not perform the leisure activities included in the E-IADL because they have no interest in doing so. Also, they may participate in leisure activities that form a separate construct, such as activities related to games or crafts, or may perform leisure activities that are scattered across multiple domains. However, these limitations do not negate the findings from this study, that is, that increasing levels of participation in sociodevelopment-type activities is associated with better cognitive function. Another limitation relates to the items themselves; some items include multiple activities in a single question, such as "attending movies, restaurants, or sporting events." Such items may have an adverse effect on construct validity. Finally, the sample is comprised of a multiethnic cohort, which supports its generality; however, this cohort is also locally restricted and urban.

Conclusions

This study constructed a reliable and valid E-IADL scale that covers a broader range of functional ability in the young–old than the traditional IADL scale. In a sample of participants aged 70, the scale dramatically reduced ceiling effects without resorting to the inclusion of physically demanding tasks, for example, heavy housework or vigorous activity; comparing the nine-item E-IADL scale with a nine-item IADL scale resulted in a reduction of ceiling effects from 67% to 3% (Figure 1). The scale met the IRT assumptions of unidimensionality, local independence, and monotonicity. The scale also met the nonparametric Rasch model equivalent of IIO, thus reinforcing construct validity by establishing a formal hierarchy of functional decline. Cross sectionally, the scale was predictive of relatively subtle cognitive differences (bottom 25% of distribution), particularly in relation to composite scores of visuospatial ability and processing speed. For both of these composites, the odds of scoring in the bottom quarter were reduced by 30% for each additional E-IADL activity endorsed. Finally,

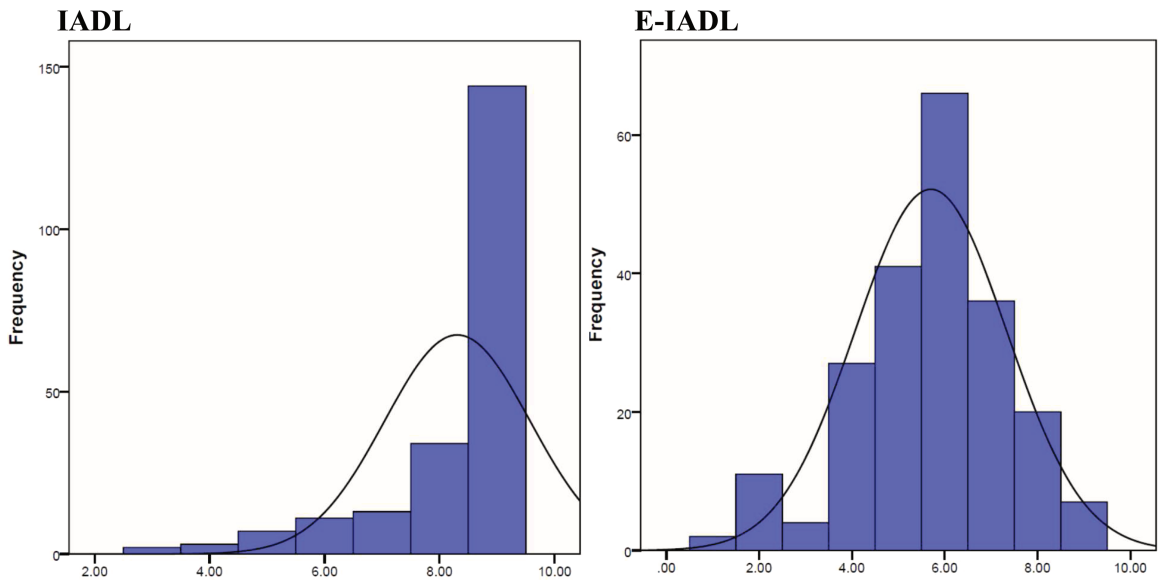

Figure 1. Sum score distributions of a traditional IADL scale and the new E-IADL scale. Note: *x*-axis represents sum scores, with a maximum score of 9 for both scales. E-IADL, extended-instrumental activities of daily living.

it is worth highlighting that these findings occur in a sample of nondemented, community-dwelling, young–old.

Supplementary Material

Supplementary material can be found at: http://biomedgerontology.oxfordjournals.org/

Funding

This research was supported by federal grant National Institutes of Health (NIH) AG07232.

References

1. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9:179–186.
2. Gold DA. An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. *J Clin Exp Neuropsychol*. 2012;34:11–34.
3. Nygård L. Instrumental activities of daily living: a stepping-stone towards Alzheimer's disease diagnosis in subjects with mild cognitive impairment? *Acta Neurol Scand, Supplc*. 2003;179:42–46.
4. Purser JL, Fillenbaum GG, Pieper CF, Wallace RB. Mild cognitive impairment and 10-year trajectories of disability in the Iowa Established Populations for Epidemiologic Studies of the Elderly cohort. *J Am Geriatr Soc*. 2005;53:1966–1972.
5. Wang CY, Hsieh CL, Olson SL, Wang CH, Sheu CF, Liang CC. Psychometric properties of the Berg Balance Scale in a community-dwelling elderly resident population in Taiwan. *J Formos Med Assoc*. 2006;105:992–1000.
6. Rajan KB, Hebert LE, Scherr PA, Mendes de Leon CF, Evans DA. Disability in basic and instrumental activities of daily living is associated with faster rate of decline in cognitive function of older adults. *J Gerontol A Biol Sci Med Sci*. 2013;68:624–630.
7. Ng TP, Niti M, Chiam PC, Kua EH. Physical and cognitive domains of the instrumental activities of daily living: validation in a multiethnic population of Asian older adults. *J Gerontol A Biol Sci Med Sci*. 2006;61:726–735.
8. Fried LP, Ettinger WH, Lind B, Newman AB, Gardin J. Physical disability in older adults: a physiological approach. Cardiovascular Health Study Research Group. *J Clin Epidemiol*. 1994;47:747–760.
9. de Morton NA, Keating JL, Davidson M. Rasch analysis of the barthel index in the assessment of hospitalized older patients after admission for an acute medical condition. *Arch Phys Med Rehabil*. 2008;89:641–647.
10. Wilms HU, Riedel-Heller SG, Angermeyer MC. Limitations in activities of daily living and instrumental activities of daily living capacity in a representative sample: disentangling dementia- and mobility-related effects. *Compr Psychiatry*. 2007;48:95–101.
11. Scarmeas N, Levy G, Tang MX, Manly J, Stern Y. Influence of leisure activity on the incidence of Alzheimer's disease. *Neurology*. 2001;57:2236–2242.
12. Wilson RS, Bennett DA, Bienias JL, et al. Cognitive activity and incident AD in a population-based sample of older persons. *Neurology*. 2002;59:1910–1914.
13. Tang MX, Cross P, Andrews H, et al. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. *Neurology*. 2001;56:49–56.
14. Van der Ark L. Mokken Scale Analysis in R. *J Stat Softw*. 2007;20(11):731–737.
15. Reeve BB, Hays RD, Bjorner JB, et al.; PROMIS Cooperative Group. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. 2007;45(5 suppl 1):S22–S31.
16. LOEVINGER J. The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychol Bull*. 1948;45:507–529.
17. van der Heijden PG, van Buuren S, Fekkes M, Radder J, Verrips E. Unidimensionality and reliability under Mokken scaling of the Dutch language version of the SF-36. *Qual Life Res*. 2003;12:189–198.
18. Roorda LD, Roebroeck ME, van Tilburg T, Lankhorst GJ, Bouter LM; Measuring Mobility Study Group. Measuring activity limitations in climbing stairs: development of a hierarchical scale for patients with lower-extremity disorders living at home. *Arch Phys Med Rehabil*. 2004;85:967–971.
19. Sijtsma K, Molenaar IW. *Introduction to Nonparametric Item Response Theory*. Thousand Oaks, CA: Sage; 2002:7–8.
20. RUMM2030 version 5.1 for Windows; RUMM Laboratory Pty Ltd, 14 Dodonaea Ct, Duncraig, WA 6023, Australia.
21. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas*. 2002;3:205–231.
22. Tennant A, Pallant JF. Unidimensionality matters. *Rasch Measurement Transactions*. 2006;20:1048–1051.
23. Moorer P, Suurmeijer TP. A study of the unidimensionality and cumulativeness of the MOS Short-form General Health Survey. *Psychol Rep*. 1994;74:467–470.
24. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. New York: McGraw Hill; 1994.
25. Bland JM, Altman DG. Cronbach's alpha (Statistics notes). *Brit Med J*. 1997;314:572.
26. Fortinsky RH, Garcia RI, Joseph Sheehan T, Madigan EA, Tullai-McGuinness S. Measuring disability in Medicare home care patients: application of Rasch modeling to the outcome and assessment information set. *Med Care*. 2003;41:601–615.
27. Dubuc N, Haley S, Ni P, Kooyoomjian J, Jette A. Function and disability in late life: comparison of the Late-Life Function and Disability Instrument to the Short-Form-36 and the London Handicap Scale. *Disabil Rehabil*. 2004;26:362–370.
28. Sikkes SA, de Lange-de Klerk ES, Pijnenburg YA, Scheltens P, Uitdehaag BM. A systematic review of instrumental activities of daily living scales in dementia: room for improvement. *J Neurol Neurosurg Psychiatr*. 2009;80:7–12.
29. Newson RS, Kemps EB. General lifestyle activities as a predictor of current cognition and cognitive change in older adults: a cross-sectional and longitudinal examination. *J Gerontol B Psychol Sci Soc Sci*. 2005;60:P113–P120.
30. Aartsen MJ, Smits CH, van Tilburg T, Knipscheer KC, Deeg DJ. Activity in older adults: cause or consequence of cognitive functioning? A longitudinal study on everyday activities and cognitive performance in older adults. *J Gerontol B Psychol Sci Soc Sci*. 2002;57:P153–P162.
31. Jopp DS, Hertzog C. Assessing adult leisure activities: an extension of a self-report activity questionnaire. *Psychol Assess*. 2010;22:108–120.
32. Kreiner S. Validity and objectivity: reflections on the role and nature of Rasch models. *Nor Psych*. 2007;59(3):268–298.
33. Balazs K, De Boeck P. Detecting local item dependence stemming from minor dimensions: Interuniversity Attraction Pole statistics network [technical report]. Retrieved from http://sites.uclouvain.be/IAP-Stat-Phase-V-VI/PhaseV/publications_2006/TR/TR0684.pdf, 2006.
34. Ligtvoet R, van der Ark LA, te Marvelde JM, Sijtsma K. Investigating an invariant item ordering for polytomously scored items. *Educ Psychol Meas*. 2010;40:578–595.
35. Sijtsma K, Junker BW. A survey of theory and methods of invariant item ordering. *Br J Math Stat Psychol*. 1996;49(Pt 1):79–105.
36. Sijtsma K, Hemker BT. Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika*. 1998;63:183–200.