

# The Adults and Older Adults Functional Assessment Inventory: A Rasch Model Analysis

Research on Aging  
2015, Vol. 37(8) 787–814  
© The Author(s) 2014  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0164027514564469  
roa.sagepub.com



Liliana B. Sousa<sup>1</sup>, Gerardo Prieto<sup>2</sup>, Manuela Vilar<sup>1</sup>,  
Horácio Firmino<sup>3</sup>, and Mário R. Simões<sup>1</sup>

## Abstract

Functional assessment methods are an important element in multi-dimensional neuropsychological evaluations, particularly in older adults. The Adults and Older Adults Functional Assessment Inventory is a new measure of basic and instrumental activities of daily living. Rasch model analyses were used to analyze the psychometric characteristics of the instrument in a sample of 803 participants. The original categories did not provide an optimal assessment of functional incapacity. The scale was dichotomized to achieve a better reliability score and item fit. The final 50 items revealed a moderately high variability in item difficulty, acceptable fits to items and persons, and a good Person Separation Reliability score. The scores were able to discriminate between normal controls and clinical patients. None of the items showed *Differential Item Functioning* associated with age, gender, or education. The instrument is able to achieve measures of functional incapacity with the useful properties of the Rasch model.

<sup>1</sup> Centro de Investigação do Núcleo de Estudos e Intervenção Cognitivo Comportamental, Faculty of Psychology and Educational Sciences, University of Coimbra, Coimbra, Portugal

<sup>2</sup> Faculty of Psychology, University of Salamanca, Salamanca, Spain

<sup>3</sup> Psychogeriatric Unit, Department of Psychiatry, Coimbra University Hospital, Portugal

## Corresponding Author:

Liliana B. Sousa, Faculdade de Psicologia e de Ciências da Educação, Universidade de Coimbra, Rua do Colégio Novo, Apartado 6153, 3001-802, Coimbra, Portugal.

Email: bapt.liliana@gmail.com

## Keywords

functional assessment, Adults and Older Adults Functional Assessment Inventory, Rasch model

## Introduction

Age projections reveal that the proportion of individuals over 60 years old is growing rapidly (World Health Organization, 2002). Projections from the Portuguese census estimate that by 2060, 32.3% of the total Portuguese population will be older than 65 years and 13.3% will be older than 80 years (Instituto Nacional de Estatística, 2009). This increase in the proportion of elderly individuals in the population has been associated with a higher prevalence of chronic conditions such as cardiovascular disease, stroke, diabetes, and musculoskeletal conditions as well as mental health conditions such as dementia and depression (World Health Organization, 2002). These conditions have a considerable incidence in adult population, although it is in the elderly population they became more relevant. These medical conditions, as well as the normative cognitive changes that occur as part of the aging process, are associated with important impairments in the capacity to perform basic activities of daily living (BADLs) and instrumental activities of daily living (IADLs; Wood et al., 2005). Therefore, the elderly population experiences a higher level of dependency and a lower quality of life (e.g., Bourdel-Marchasson, Helmer, Fagot-Campagna, Dehail, & Joseph, 2007; Wada et al., 2005).

In 1990, the World Health Organization introduced the term “Active Aging,” which was defined as “... the process of optimizing opportunities for health, participation and security in order to enhance quality of life as people age” (World Health Organization, 2002, p. 12). The year 2012 was the international year of Active Aging, in which the fundamental goal was to maintain and/or increase the quality of life in older adults as well as to ensure their autonomy and independence. Autonomy refers to the ability to make decisions according to one’s own values and preferences, and independence refers to the ability to perform the BADLs and IADLs with little or no help from others (World Health Organization, 2002). To help an individual achieve autonomy and independence in clinical practice, a multidimensional evaluation of the individual must include an assessment of that individual’s ability to conduct BADLs and IADLs (e.g., Burns, Lawlor, & Craig, 2004; Potter & Attix, 2006).

This functional assessment should be integrated with neuropsychological evaluations for several reasons. First, functional limitations are considered strong predictors of overall health (Marengoni et al., 2004) and death as well as of the likelihood of an admission to a nursing home (Gill, 2010). Furthermore, although cognitive function and daily living function are closely associated, neuropsychological tests of cognitive function do not explain all of the variance in the ability to perform daily living activities (Potter & Attix, 2006). In addition, functional decline is also used as a separate criterion from cognitive impairment to identify neurodegenerative conditions such as mild cognitive impairment and dementia (Marson & Hebert, 2006).

### *Functional Assessment*

Functional capacity encompasses a wide range of specific abilities that are required to function independently in daily living. The BADLs refers to self-care tasks, including feeding, dressing, bathing, continence, mobility, and transference. These activities normally involve automatic procedural memory processes and basic motor functions but do not require attentional processes. In contrast, IADLs requires higher level cognitive functions (memory, attention, and executive function) and refers to complex tasks that are necessary to function independently in the home and the community. These IADLs include the preparation of meals, housekeeping tasks, and home security (IADLs-*Household*; IADLs-H) as well as comprehension and communication skills to make medical or financial decisions (IADLs-*Advanced*; IADLs-A; Marson & Hebert, 2006). There are several existing methods to determine functional capacity. Although self-report measures and reports by a third party (e.g., a family member or caretaker) are the most common methods in clinical practice, direct observation and performance-based methods show greater benefits than self-report measures (Moore, Palmer, Patterson, & Jeste, 2007). However, some studies have concluded that there are no significant differences between these methods (Hoeymans, Feskens, van den Bos, & Kromhout, 1996; West et al., 1997), meaning that the association between performance and self-report is strong.

Several instruments are available to examine functional status. Some instruments have been designed specifically to examine the BADLs, including the Katz (Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963) and Barthel Indexes (Mahoney & Barthel, 1965). Other instruments, such as the Lawton and Brody Instrumental Activities of Daily Living Scale (Lawton & Brody, 1969), have been designed to assess the IADLs. However, some instruments include items to assess both BADLs and IADLs tasks, for example the

Functional Independence Measure (FIM; Keith, Granger, Hamilton, & Sherwin, 1987) and the Functional Activities Questionnaire (FAQ; Pfeffer, Kurosaki, Harrah, Chance, & Filos, 1982). In addition, several instruments have been developed to assess functional capacity in specific medical conditions, such as the Disability Assessment for Dementia Scale (DAD; Gélinas, Gauthier, McIntyre, & Gauthier, 1999) for dementia, the Alzheimer's Disease Cooperative Study Scale for ADL in Mild Cognitive Impairment (ADCS MCI ADL; Galasko et al., 1997) for mild cognitive impairment, and the World Health Organization Disability Assessment Schedule-II (WHODAS-II; World Health Organization, 2000) for general mental health conditions.

These functional assessment instruments use several methods to determine functional capacity, including by difficulty level (e.g., the WHODAS-II includes four difficulty levels, namely, "mild," "moderate," "severe," and "extreme"), by dependence/independence level (e.g., the Barthel Index includes options for "dependent" or "independent" as well as "needs help" for a subset of items), and by execution level (e.g., the laundry item of the Lawton & Brody scale includes three execution levels, that is, "does personal laundry completely," "launders small items; rinses stockings, etc.," and "all laundry must be done by others"). These methods offer distinct ways to assess functional capacity. However, this variety of methods also hinders the ability to develop a comprehensive and integrative functional assessment instrument. Therefore, an Item Response Theory (IRT) procedure such as the Rasch model may provide an optimal way to develop new functional measures.

### *Applying the Rasch Model to Functional Assessment Instruments*

In Classical Test Theory, interpretations are based on group-referenced norms. The main advantage of the Rasch model is that this model considers interactions between persons and items using the same *logits* interval scale for both (Hobart & Cano, 2009). Therefore, the Rasch model facilitates the interpretation of the relationship between latent variables and items (Thomas, 2011). Although Rasch models were originally developed for the analysis of dichotomous items (two response categories), these models have been adapted to analyze polytomous items (more than two response categories). For dichotomous items, Rasch model is represented as:

$$P_{ni} = \exp(B_n - D_i) / (1 + \exp[B_n - D_i]),$$

where

$P_{ni}$  is the probability that person  $n$  passes item  $i$ ,

$B_n$  is the person ability level, and  
 $D_i$  is the item location (Rasch, 1960).

For the polytomous items, the Rating Scale Model (RSM) is the more used model (Andrich, 1978). The RSM, which is an extension of the Rasch model (Thomas, 2011) with good metric properties (Prieto & Delgado, 2007), is represented as:

$$\log(P_{nik}/P_{ni(k-1)}) = B_n - D_i - F_k,$$

where

$P_{nik}$  is the probability that person  $n$  chooses category  $k$  for the response to item  $i$ ,

$P_{ni(k-1)}$  is the probability that person  $n$  chooses category  $k - 1$  for the response to item  $i$ ,

$B_n$  is the overall ability level of person  $n$ ,

$D_i$  is the difficulty of item  $i$ , and

$F_k$  is the likelihood of choosing a response from category  $k$  relative to  $k-1$ .

This step calibration is a rating scale threshold that is defined as the location that corresponds to an equal probability of a response in adjacent categories  $k-1$  and  $k$  (Andrich, 1978; Bond & Fox, 2007).

The RSM has been important for instrument development (Walker, Böhnke, Cerny, & Strasser, 2010) because it enables an empirical study of the response categories (Knutsson, Rydstrom, Reimer, Nyberg, & Hagell, 2010). An analysis of the response categories is important for instrument development because response categories must reflect the construct to be assessed and should not produce ambiguous responses (Bond & Fox, 2007). The RSM has been applied to several functional assessment instruments (Lindeboom, Vermeulen, Holman, & de Haan, 2003). Rasch models and other IRT procedures have also been useful in the psychometric characterization of functional assessment measures, including reliability, construct validity, content validity (Fieo, Austin, Starr, & Deary, 2011), and dimensionality (Breithaupt & McDowell, 2001). The psychometric characteristics of several well-known functional assessment instruments, including the Lawton IADL scale (McGrory, Shenkin, Austin, & Starr, 2013), Barthel Index (Morton, Keating, & Davidson, 2008), and the FIM (Granger, Deutsch, & Linn, 1998), have been assessed by IRT procedures.

The main purpose of this study is to apply the RSM to the Adults and Older Adults Functional Assessment Instrument—experimental version

**Table 1.** Sample Characteristics.

		Comparison Group (n = 567)	Clinical Group (n = 236)
Gender	Men	186 (33%)	74 (31%)
	Woman	381 (67%)	162 (68%)
Age	<60	32 (7%)	65 (28%)
	60–64	144 (25%)	27 (12%)
	65–69	110 (19%)	33 (14%)
	70–74	113 (20%)	40 (17%)
	75–79	85 (15%)	43 (18%)
	80–84	65 (11%)	15 (6%)
	>84	18 (3%)	13 (5%)
	<i>M</i> ± <i>SD</i>	69.92 ± 7.87	65.35 ± 14.72
Education <sup>a</sup>	0–2	53 (9%)	30 (13%)
	3–4	327 (58%)	123 (52%)
	5–9	108 (19%)	29 (12%)
	10–12	36 (6%)	23 (10%)
	> 12	43 (8%)	13 (5%)

Note. *M* = Mean; *SD* = Standard Deviation; *N* = 803 participants.<sup>a</sup>18 missing values in clinical group.

(*Inventário de Avaliação Funcional de Adultos e Idosos* [IAFAI]; Sousa, Simões, Pires, Vilar, & Freitas, 2008), which is a new instrument to assess the functional incapacity of adults and older adults that includes BADL, IADL-H, and IADL-A items. The RSM is used to study the original response categories of the IAFAI as well as its dimensionality, reliability, item difficulty, fit indexes for items and persons, ability to differentiate normal controls from patients with several clinical conditions, effects of age, gender, and education as well as Differential Item Functioning (DIF).

## Method

### *Participants and Procedures*

The sample of 803 participants (Table 1) included a comparison group of 567 community-dwelling adults and older adults and a clinical group of 236 patients with several neurological (mild cognitive impairment, dementia, epilepsy, traumatic brain injury, and stroke) or psychiatric diagnoses (depression, anxiety, and schizophrenia). Participants were excluded from the comparison group if they had an actual or previous neurological, psychiatric, or psychological disease as well as if they had some orthopedic or other medical

condition that affects functional status. Informed consent was obtained from all participants. A trained psychologist administered the neuropsychological assessment to each participant that includes not only the IAFAI but also the instruments for cognitive and depressive symptoms screening. Participants in the clinical group were referred to the study by their doctors and examined in a hospital setting. Only the participants with recognized diagnosis were considered. Participants in the comparison group were assessed in the community (through the presentation of the study in day care centers and parish councils) or in general medical centers (in contexts of the routine medical examinations), all over the country.

### *Measure*

The IAFAI (Sousa et al., 2008) is a new comprehensive instrument to assess the functional incapacity of adults and older adults. The IAFAI was developed to provide a useful and specialized tool to be used in contexts of neuropsychological assessment in both clinical and forensic settings (Sousa, Simões, Firmino, & Peisah, 2013). During the development of the IAFAI, the conceptual model considered was the model proposed by Marson and Hebert (2006) who differentiated daily living activities into three main groups: BADL, IADL-H, and IADL-A (Marson & Hebert, 2006). The International Classification of Functioning, Disability, and Health (World Health Organization, 2001) was also considered to integrate contextual factors in the definition of functional incapacity.

The IAFAI includes both BADL and IADL items to enable a comprehensive assessment of functional incapacity in neuropsychological evaluations. A prior study has shown that including both BADL and IADL items in the same scale improves measurement sensitivity (Spector & Fleishman, 1998). The first experimental version of the IAFAI was composed of 84 items; however, the final experimental version (Sousa, Vilar, Pires, Freitas, & Simões, in press) that will be studied with IRT analysis is composed of 53 items, including 18 BADL items, 18 IADL-H items, and 17 IADL-A items. The BADL items encompass four domains (feeding, dressing, bathing and continence, and mobility and transference), the IADL-H items encompass four domains (conversation and telephone use, meal preparation, housekeeping, and home security), and the IADL-A items encompass five domains (comprehension and communication, health-related decision making, finances, going out and transportation use, and leisure and interpersonal relationships).

Existing functional assessment instruments use several methods to determine functional incapacity, including by difficulty level, by dependence/

independence level, or by execution level. The IAFAI attempts to combine these measurement approaches to generate a more reliable indicator of functional incapacity. Nine distinct response categories were developed for the IAFAI to assess the dynamic process of functional decline, namely, the independence levels include “independence without difficulty,” “independence with little difficulty,” “independence with much difficulty,” and “modified independence” (i.e., independence that involves some external devices); the dependence levels include “supervision without difficulty,” “supervision with little difficulty or help without difficulty,” “supervision with much difficulty or help with little difficulty,” “help with much difficulty,” and “incapable/unable to do” (i.e., the extreme level of functional incapacity; Sousa et al., in press).

### *Statistical Analysis*

Descriptive statistics were computed with the Statistical Package for Social Sciences 20.0 (SPSS 20.0; IBM SPSS, Chicago, IL). Rasch analyses were performed in WINSTEPS (Linacre, 2012). The RSM was used because all of the IAFAI items include multiple response categories (Wright, 1999). To analyze the functionality of the response categories, the Linacre (2002) criteria were used, that is, (i) a minimum of 10 observations from each response category, (ii) a regular distribution of the observations among the categories, (iii) a monotonic increase in the average measure in each category, (iv) an average residual (infit and/or outfit) with a value less than 2.0, and (v) a monotonic increase in the step calibration between categories. When some of these criteria are not met, adjacent categories should be combined, and the data should be reanalyzed (Andrich, de Jong, & Sheridan, 1997). The IAFAI category responses were also analyzed visually with category characteristic curves. After the category response analysis was conducted, the model fit was analyzed for persons and items. In our study, fit analysis was done using outfit and infit indexes. Outfit is the mean of the squared standardized residuals (differences between the observed responses and those predicted by the model) and infit is the mean of the squared standardized residuals weighted by the information function. The interpretation of the misfit values (infit and/or outfit) followed the criteria that were established by Linacre (2012), that is, (i) values between 0.5 and 1.5 indicate that the items are important for the measure, (ii) values between 1.5 and 2.0 indicate that the items produce a moderate misfit to the measure, and (iii) values higher than 2.0 indicate that the items produce a severe misfit to the measure (and should be excluded from the measure).



Because Rasch models are highly dependent on unidimensionality (Tennant & Pallant, 2006), the dimensionality of the IAFAI was analyzed using a principal component analysis (PCA) of the residuals. This analysis looks for patterns in the residuals, which represent the portion of the data that do not agree with the Rasch measures. PCAs attempt to find a component that explains the largest amount of variance in the residuals under the assumption that the residuals do not represent random noise. Linacre (2012) proposes that a fundamental unidimensionality exists if the eigenvalue of the first component of the residuals is small (usually less than 2.0) and the percentage of the raw explained variance is large (usually over 50% as a rule thumb).

Significant contrasts between normal control and clinical group means were performed using Welch's  $t$ , which is an adaptation of Student's  $t$ -test intended for use with two samples having possible unequal variances. This same test was used to explore the effect of age, gender, and education on functional incapacity score measured by IAFAI. The ability of the IAFAI items to discriminate between the normal controls and the clinical group was done through the probability difference between both the groups ( $P_N - P_C$ ;  $P_N$  is the probability of a person with a mean ability of the control group,  $P_C$  is the probability of a person with a mean ability of the clinical group).

The most important property of the Rasch model, known as *specific objectivity* (Andrich, 1988), means that individuals with the same ability ( $B$ ) will have the same likelihood of correctly answering an item, regardless of whether they belong to groups with different cultures, gender, or native language. The DIF detection procedure in the Rasch model is based on the item characteristic curve (ICC), the proportion of individuals at the same ability level who answer a given item correctly. If the item measures the same ability across groups then, except for random variations, the same proportion is found irrespective of the nature of the group, that is, in the absence of DIF, the ICC in the different groups and the item parameter of difficulty ( $D$ ) will be invariant. Thus, the hypothesis of the absence of DIF was tested by calculating the difference between the estimators of the item parameter of difficulty for each group ( $D_f - D_r$ ), thus controlling for the possible differences between the groups (focal and reference) in the latent variable. Wright and Douglas (1976) found that differences lower than 0.50 logits had negligible consequences regarding the validity of the measure. The  $t$ -test with the Bonferroni adjustment (Benjamini & Hochberg, 1995) was used to test the significance and is described as:

$$t = D_f - D_r / (SE_{D_f}^2 + SE_{D_r}^2)^{1/2},$$

where

$D_f$  is the difficulty parameter in focal group,  
 $D_r$  is the difficulty parameter in reference group,  
 $SE_{Df}$  is the standard error of difficulty parameter in focal group, and  
 $SE_{Dr}$  is the standard error of difficulty parameter in reference group.

According to this method, the presence of DIF is detected by a difference greater than 0.50 logits and statistically significant (Bonferroni's correction:  $p = .05/50 = .001$ ) between the difficulty parameters of the reference group and the focal group.

The Mantel–Haenszel (MH) method was also used for DIF analysis. The procedure is based on an analysis of the contingency tables corresponding to the different levels in which the variable has been divided. For each level  $j$ , the odds ratio ( $\alpha$ ) is calculated as:

$$\alpha = \left( \frac{p_{Rj}}{1 - p_{Rj}} \right) / \left( \frac{p_{Fj}}{1 - p_{Fj}} \right),$$

where

$p_{Rj}$  is the odd of a correct answer to the item in the reference group, and  
 $p_{Fj}$  is the odd of a correct answer to the item in the focal group.

The null hypothesis of the absence of DIF can be tested using the chi-square statistic ( $MH\chi^2$ ; Holland & Thayer, 1988), which is distributed as  $\chi^2$  with one degree of freedom. Testing the absence of DIF on a test involves multiple comparisons (at least 1 for each item). Zwick and Ercikan (1989) found that differences lower than 1.5 Delta-MH (0.64 logits) had negligible consequences as regards the validity of the measures. Thus, the DIF is usually considered substantial if the Delta MH value is classified as C ("large DIF," according to the criteria of the Educational Testing Service), that is, size higher than 0.64 and significant  $\chi^2$  statistic (using Bonferroni's correction).

## Results

### Participants

The main demographic characteristics of the sample are presented in Table 1, for the comparison and clinical groups. There are a higher percentage of women in both groups. The mean age is also equivalent in both the groups (comparison group:  $M = 69.92$ ;  $SD = 7.87$ ; clinical group:  $M = 65.35$ ;

**Table 2.** IAFAI: Analysis of the Categories' Properties.

	Chosen <i>F</i>	%	Average <i>B</i>	Infit	Outfit	Step
Category (1)						
0	30,066	81	-1.33	1.14	1.15	None
1	3,154	08	-.75	1.10	.84	1.06
2	1,588	04	-.45	.88	.59	.01
3	323	01	-.48	1.06	1.50	1.11
4	174	00	-.42	1.22	1.52	.26
5	442	01	-.20	.94	.76	-1.19
6	178	00	-.11	.83	.92	.73
7	171	00	-.08	1.00	1.33	-.06
8	1,122	03	-.06	1.17	1.55	-1.92
Category (2)						
0	30,066	81	-2.65	1.10	1.12	None
1	5,065	14	-1.11	.93	.72	-.27
2	2,087	06	-.19	1.09	1.11	.27

Note. IAFAI = Inventário de Avaliação Funcional de Adultos e Idosos; Chosen *F*(%) = Observed count and percentage of occurrences in each category; average *B* = The average of the measures that are modeled to produce the responses observed in each category; infit/outfit = The average of the *infit* and *outfit* mean squares associated with the responses in each category; step = Rating scale threshold between two adjacent categories *K* and *K* - 1. (1) Original categories; (2) Modified categories (Category 0 = original 0 category; Category 1 = 1, 2, and 3 original categories; Category 2 = 4, 5, 6, 7, and 8 original categories).

*SD* = 14.72). A quite higher proportion of the sample has fewer years of formal education (4 years or less).

### Response Categories

The original nine categories of the IAFAI do not adequately assess functional incapacity because the category thresholds are disordered. In addition, the average measure and the step calibrations by category do not change monotonically (Table 2). Therefore, the original categories were collapsed into three categories to obtain a better assessment of functional incapacity. However, the three modified categories also failed to produce an optimal assessment of functional incapacity because the central category appears to have less utility than the other two categories due to a compressed range of responses. Therefore, the analysis was repeated for two categories (0 = *total independence category*; 1 = *modified independence/dependence category*). According to these categories, functional incapacity in performing some

daily living activity represents not only the dependence on others but also the difficulty in performing that daily living activity. By doing this, we try to detect the minor changes in functional capacity with IAFAI scores. These two categories result in a better item fit and fewer items with a moderate misfit (4 misfit items) compared with the model with three categories (8 misfit items). In addition, the person variability and the person separation reliability (PSR) scores are both higher in the model with two categories (PSR = 0.79) compared to the model with three categories (PSR = 0.75).

### *Dimensionality*

The Rasch PCA of the residuals was conducted on the dichotomized items and shows that the percentage of the raw variance that is explained by the Rasch measures is higher (34.0%) than the minimum acceptable value for unidimensionality (20%) proposed by Reckase (1979). The PCA of the residuals also shows no discernible pattern (the first factor explains only 5.7% of the variance in the residuals), which further supports unidimensionality (Tenant & Pallant, 2006). The flexible consideration of the unidimensionality of the IAFAI is also supported by the negligible number of moderate misfit items and no severe misfit items that emerged from the analysis.

### *Fit Indexes for Items and Persons*

The next analyses were conducted on dichotomized items. Three IAFAI items were eliminated (*Using a computer*, *Driving near your area of residence*, and *Driving far from your area of residence*) because these items had a moderate misfit (1.90, 1.98, and 1.98, respectively) and did not apply to the majority of the sample population. The item using a computer was answered by only 168 of the total 803 participants (21% of the total sample). The items driving near your area of residence and driving far from your area of residence were only applicable to 322 (40%) and 314 (39%) subjects, respectively, in the sample population. Once these items were excluded and the optimal number of categories was established, new data analyses were conducted to quantify the model fit and the indicators of validity as well as to determine the item and person parameters (Tables 3 and 4). The final 50 items of the IAFAI are presented in the Appendix.

The item difficulty ranges from  $-2.05$  (Item 10) to  $1.95$  logits (item 15; Figure 1). The item difficulty parameters are estimated with a good reliability ( $SE: M = 0.12$ ;  $SD = 0.02$ ; item separation reliability = .98), which means that the IAFAI items were measured with a high precision. The classical

**Table 3.** IFAI: Statistics of the Items.

Item	<i>P</i>	RiX	D	SE	Infit	Outfit	$P_N$	$P_C$	$P_N - P_C$
1	.08	.37	1.22	.15	1.07	.98	.02	.06	-.04
2	.09	.46	1.20	.14	.85	.56	.02	.06	-.04
3	.10	.47	1.00	.14	.88	.67	.02	.07	-.05
4	.10	.51	1.00	.14	.78	.44	.02	.07	-.05
5	.21	.60	-.28	.11	.84	.68	.08	.22	-.14
6	.17	.57	.12	.11	.82	.61	.06	.16	-.10
7	.13	.52	.55	.12	.87	.81	.04	.11	-.07
8	.33	.60	-1.32	.10	1.01	.94	.21	.44	-.23
9	.18	.62	.07	.11	.70	.43	.06	.16	-.10
10	.05	.39	1.95	.19	.83	.38	.01	.03	-.02
11	.10	.48	.95	.14	.88	.60	.03	.07	-.04
12	.26	.49	-.69	.10	1.21	1.22	.12	.30	-.18
13	.11	.50	.84	.13	.83	.65	.03	.08	-.05
14	.27	.52	-.75	.10	1.12	1.25	.13	.31	-.18
15	.45	.56	-2.05	.09	1.16	1.25	.35	.62	-.27
16	.31	.57	-1.09	.10	1.04	.99	.17	.38	-.21
17	.23	.58	-.38	.10	.93	.73	.09	.24	-.15
18	.20	.57	-.20	.11	.91	.74	.07	.21	-.14
19	.14	.55	.38	.12	.79	.64	.05	.12	-.07
20	.06	.44	1.58	.17	.80	.55	.01	.04	-.03
21	.30	.53	-.98	.10	1.13	1.13	.15	.36	-.21
22	.21	.51	-.23	.11	1.03	1.13	.08	.21	-.13
23	.14	.52	.48	.12	.91	.66	.04	.11	-.07
24	.08	.51	1.07	.15	.72	.30	.02	.07	-.05
25	.18	.58	-.11	.13	.80	.86	.07	.16	-.09
26	.14	.57	.35	.14	.76	.49	.05	.13	-.08
27	.27	.65	-.79	.12	.75	.61	.13	.31	-.18
28	.40	.62	-1.79	.10	.97	.93	.29	.53	-.24
29	.15	.57	.17	.13	.82	.58	.06	.15	-.09
30	.19	.59	-.15	.13	.83	.66	.07	.19	-.12
31	.12	.41	.58	.14	1.04	1.19	.04	.11	-.07
32	.40	.41	-1.70	.09	1.49	1.98	.27	.53	-.26
33	.16	.43	-.07	.12	1.12	1.35	.06	.18	-.12
34	.09	.39	1.07	.15	1.02	1.11	.02	.07	-.05
35	.08	.39	1.17	.16	.99	1.04	.02	.06	-.04
36	.06	.40	1.34	.17	.87	.49	.02	.06	-.04
37	.33	.61	-1.20	.10	.96	.93	.19	.41	-.22
38	.29	.57	-.96	.10	1.01	1.05	.15	.35	-.20
39	.18	.56	-.10	.12	.93	.67	.07	.19	-.12
40	.19	.53	-.11	.11	.99	.82	.07	.19	-.12

(continued)

Table 3. (continued)

Item	<i>P</i>	RiX	D	SE	Infit	Outfit	$P_N$	$P_C$	$P_N - P_C$
41	.22	.50	-.43	.11	1.07	1.20	.10	.24	-.14
42	.18	.53	-.02	.11	.96	.77	.06	.17	-.11
43	.30	.60	-1.45	.12	.98	1.04	.22	.47	<b>-.25</b>
44	.17	.37	-.11	.15	1.43	1.60	.07	.19	-.12
45	.18	.39	-.02	.13	1.38	1.28	.06	.17	-.11
46	.21	.39	-.32	.13	1.42	1.62	.09	.22	-.13
47	.09	.24	1.23	.16	1.45	1.53	.02	.06	-.04
48	.18	.43	.07	.13	1.27	1.18	.06	.17	-.11
49	.23	.51	-.44	.12	1.19	.98	.10	.24	-.14
50	.26	.43	-.63	.11	1.33	1.44	.11	.28	-.17

Note. IAFAI = Inventário de Avaliação Funcional de Adultos e Idosos;  $p$  = proportion of persons that have functional incapacity (Score 1) in performing the daily living activity; RiX = Item-total correlations; D = Difficulty of the items; SE = standard error; infit/outfit = Rasch model adjustment parameters;  $P_N$  = probability of not doing the daily living activity for  $B = -2.67$  (mean of the normal control group);  $P_C$  = probability of not doing the daily living activity for  $B = -1.56$  (mean of the Clinical group);  $P_N - P_C$  = probability difference between normal control and clinical groups (discriminant efficiency; highest presented in boldface).

Table 4. IAFAI: Summary of the Statistics for Items and Persons.

Statistics	Item Statistics						Person Statistics		
	$p$	RiX	D	SE	Infit	Outfit	X	B	SE
Mean	.19	.50	.00	.12	1.00	.92	8.60	-2.34	.71
Standard deviation	.10	.09	.92	.02	.20	.36	9.20	1.93	.49
Maximum	.45	.65	1.95	.19	1.49	1.98	46.00	3.88	1.84
Minimum	.05	.24	-2.05	.09	.70	.30	0.00	-5.52	.31
% Moderate misfit (1.5-2.0)	—	—	—	—	0%	8%	—	—	—
% Severe misfit (>2.0)	—	—	—	—	0%	0%	2%	—	—
			ISR = .98				PSR = .79 Cronbach's $\alpha = .93$		

Note.  $p$  = proportion of persons who have functional incapacity (Score 1) in performing the daily living activity; RiX = item-total correlations; D = difficulty of the items; SE = standard error; infit/outfit = Rasch model adjustment parameters; X = Number of activities of daily living where the subjects have functional incapacity (Score 1); B = ability of the persons; ISR = item separation reliability; PSR = person separation reliability.

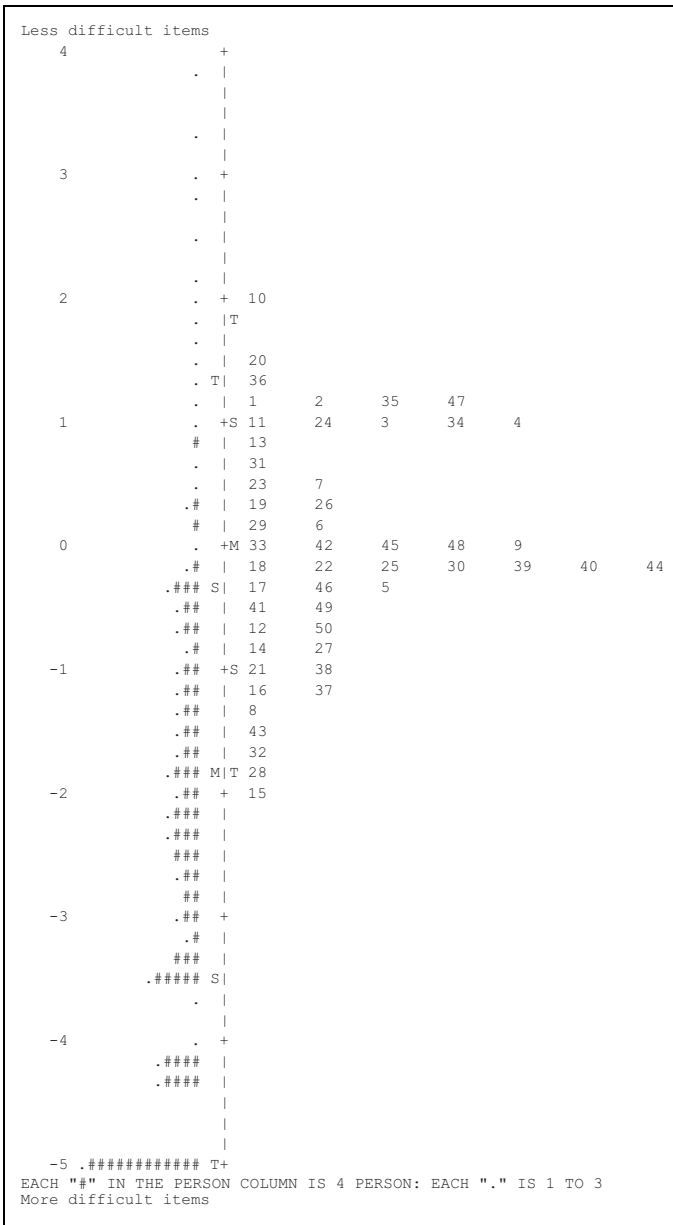


Figure 1. Inventário de Avaliação Funcional de Adultos e Idosos (IAFAI): item-person map. Each “#” in the person column is 4 persons and each “.” is 1–3.

difficulty index was also computed ( $p$ ) to assess the difficulty of the items. In this study,  $p$  represents the proportion of individuals who have functional incapacity (Score 1) in performing the daily living activity for each item. The minimum  $p$  value was .05 (Item 10) and the maximum  $p$  value was .45 (Item 15). The mean  $p$  value was .19, which indicates that 19% of the individuals in the sample population have functional incapacity (Score 1) in performing the daily living activity in each items evaluated by IAFAI ( $p$ :  $M = 0.19$ ;  $SD = 0.10$ ). The low values of the difficulty indexes are associated with the sample characteristics because the majority of the sample includes normal controls. This is also the main reason for some of the floor effect detected.

The item-total correlations are moderately high (RiX:  $M = 0.50$ ;  $SD = 0.09$ ) with values between 0.24 (Item 47) and 0.65 (Item 27). The item fit to the model is acceptable, with none of the items resulting in a severe misfit (i.e., an outfit and/or infit values higher than 2.0). A moderate misfit (i.e., outfit and/or infit values between 1.5 and 2.0) occurred in only 4 (8%) of the 50 items. Although these items revealed a moderate misfit, they were applied to a large proportion of the sample and were not excluded from the inventory.

Table 4 presents the main statistical results about the person fit statistics. In the analyzed sample ( $n = 567$  normal controls, 70.6% of the total sample, and  $n = 236$  clinical patients, 29.4% of the sample), the mean number of daily living activities where the participants have functional incapacity (Score 1) is low ( $X$ :  $M = 8.60$ ;  $SD = 9.20$ ). However, the variability in the levels of functional incapacity is high, with a range between 0 and 46. These results are similar to the results that were observed on the logit scale, in which the sample mean indicated low levels of functional incapacity but high variability in the scores ( $B$ :  $M = -2.34$ ;  $SD = 1.93$ ; values between  $-5.52$  and 3.88). These results may be attributed to the composition of the sample population, which includes mostly normal control participants. The PSR (PSR = 0.79) score is acceptable and is associated with a high Cronbach's  $\alpha$  ( $\alpha = .93$ ). There are a negligible number of individuals with a severe misfit (only 2% of the total sample).

### *Ability to Discriminate Normal Controls From Clinical Conditions*

The IAFAI scores are able to discriminate between the comparison group ( $M = -2.67$ ) and the clinical group ( $M = -1.56$ ). The difference in the means between the two groups ( $-1.12$ ) is statistically significant ( $t = -7.61$ ;  $p < .01$ ) and is associated with a medium effect size (Cohen's  $d = 0.60$ ). The results using the Rasch logistic function equation (Table 3) reveal that some of the IAFAI items are significantly different between the normal controls



and the clinical patients. Some of the highest discriminative items (presented in boldface) include items 15, 32, and 43.

### *Effects of Age, Gender, and Education*

The IAFAI scores are higher in older ( $\geq 70$  years old;  $n = 392$ ;  $M = -1.97$ ) than in younger ( $< 70$  years old;  $n = 411$ ;  $M = -2.70$ ) persons, meaning that higher age is associated with higher levels of functional incapacity. This difference is statistically significant ( $t = -5.43$ ;  $p < .01$ ) but is associated with a small effect size ( $d = 0.39$ ). Concerning gender, males ( $n = 260$ ;  $M = -2.58$ ) have better functional status than females ( $n = 543$ ;  $M = -2.23$ ). Despite the significant difference ( $t = -2.34$ ;  $p = .019$ ), the effect size was small ( $d = 0.18$ ). The less educated persons ( $\leq 4$  years;  $n = 533$ ;  $M = -1.94$ ) have higher scores (poorer functional status) than the persons with higher education levels ( $> 4$  years;  $n = 252$ ;  $M = -3.33$ ). This is a statistically significant difference ( $t = 10.50$ ;  $p < .01$ ) and associated with a medium effect ( $d = .79$ ).

### *DIF*

DIF analyses were conducted to explore the likelihood that individual items of the IAFAI might work differently as a function of group (normal controls vs. clinical), age ( $< 70$  years vs.  $\geq 70$  years), gender (male vs. female), and educational level ( $\leq 4$  vs.  $> 4$ ). The absence of DIF involves a difference lower than 0.50 logits (without statistical significance) between the estimators of the item parameter of difficulty for each group and a Delta MH value classified as C (i.e., size higher than 0.64 and significant). The results reveal that there are no items of the IAFAI with DIF associated with age, gender, and education. Two items (Number 8 and 15) revealed DIF associated with group being more difficult for the clinical group.

## **Discussion**

An empirical study of the original response categories on the IAFAI was performed with the RSM, which is an extension of the Rasch model that accounts for polytomous items (Bond & Fox, 2007; Wright, 1999). The results did not support the original nine categories on the IAFAI. Therefore, the categories were consolidated by collapsing adjacent categories. The best model included only two categories, with a score of 0 representing the absence of functional incapacity (absence of difficulty or dependence in the execution of the activity of daily living) and a score of 1 representing the presence of functional

incapacity (presence of difficulty or dependence in the execution of the activity of daily living). The Rasch analysis of the dichotomized IAFAI items reveals a better reliability and item fit parameters. Despite the existence of different and distinct ways to measure function (difficulty level, dependence/independence level, execution level), which were integrated with IAFAI initial categories, the results demonstrate that more simple methods are preferable. The dichotomous categories not only made an instrument more easily administered but also improved its psychometric characteristics.

Finlayson, Mallinson, and Barbosa (2005) also found that a dichotomous rating scale provided a better fit (with no misfit items and a higher person variability) on the AIM Longitudinal Study compared to a rating scale with five response categories in a sample of 607 older adults (238 living at home without services, 187 living at home with some care services, and 182 living in a nursing home). In addition, a study of the Motor Subscale of the FIM showed that seven categories provided an adequate fit for only 5 of the 13 items; for the remaining items, dichotomous categories provided a better fit (Tennant et al., 2004). Dichotomous categories have been used in several functional assessment instruments, including those measuring BADL (e.g., Katz Index) and IADL (e.g., Disability Assessment for Dementia Scale; DAD). However, several functional assessment instruments use more than two categories—three categories (e.g., the ADL subscale of the Older Americans Resources and Services Program; OARS); four categories (e.g., WHO-DAS-II); five categories (e.g., Health Assessment Questionnaire); and seven categories (FIM). Despite this evidence, the majority of these instruments have not been analyzed with IRT procedures or the Rasch model. Studies in which these analyses have been conducted have concluded that category reduction is necessary (e.g., the Tennant et al., 2004 study of the FIM) and is associated with an improvement in the psychometric characteristics of the instruments (Tennant et al., 2004).

The analysis in this study revealed the IAFAI essential unidimensionality, which agrees with other studies (Spector & Fleishman, 1998; Finlayson, Mallinson, & Barbosa, 2005; LaPlante, 2010). However, the dimensionality of the items that evaluate the BADLs and IADLs is inconsistent across studies. For example, Breithaupt and McDowell (2001) found a two-factor structure in which ADLs and IADLs items represented different dimensions that were strongly correlated ( $r = .79$ ). Thomas, Rockwood, and McDowell (1998) found a factor structure with three main factors that were related to “basic self-care,” “medium self-care,” and “complex management.”

In this study, both a low item difficulty value and a low mean value of functional incapacity were found. These values are associated with the

distribution of the sample population and indicate a higher number of normal controls than clinical patients in the sample. Other studies have also found that normal controls report greater independence compared with dementia patients (Breithaupt & McDowell, 2001). Similar results have been found in samples that are composed of individuals with several clinical conditions (Morton et al., 2008). Some studies have attempted to determine the point at which the older population begins to experience functional limitations. Community-dwelling older adults appear to start to lose the ability to perform the more complex activities of daily living around age 80 (Royall et al., 2007). This result may explain the lower values of functional incapacity that are observed in community-dwelling older adults when lower age-groups are included in the sample.

I-AFAI scores were able to discriminate between comparison group and clinical patients. The items that were associated with the greatest ability to discriminate between these two groups include Item 15 (BADL item), Item 32 (IADL item), and Item 43 (IADL item). Breithaupt and McDowell (2001) found that BADLs items (Getting out of bed, Toilet transfer, and Dressing), and IADLs items (Shopping, Getting places, and Preparing meals) were the best discriminators between dementia patients and normal controls in a sample of 1,364 elderly Canadians from the Canadian Study of Health and Aging (Breithaupt & McDowell, 2001). These results are not directly comparable to the present study because the Canadian Study of Health and Aging considered a specific clinical group (Dementia) instead of a more general clinical population. However, the ability to discriminate between comparison group and a clinical group by I-AFAI scores agrees with several studies that have found a decline in functional capacity in distinct clinical conditions, including depression (Wada et al., 2005), schizophrenia (Green, Kern, & Heaton, 2004), mild cognitive impairment (Yeh et al., 2011), dementia (Sauvaget, Yamada, Fujiwara, Sasaki, & Mimori, 2002), and stroke (Landi et al., 2006). Although these clinical conditions have been aggregated in this study, each condition has been associated with functional decline in previous studies. Additional analyses revealed that I-AFAI scores are associated with age, gender, and education—poorer functional status was observed in older persons, females, and with lower educational levels. This was also observed in previous studies (Østerås et al., 2007; Palacios-Ceña et al., 2012).

DIF analyses showed that the items have invariance properties for young and older adults, males and females, lower and higher educated individuals, as no items showed age-, gender-, and education-related DIF. Only 2 items revealed DIF associated with group. According to this, the I-AFAI scores are able to measure the same level of functional incapacity in young and older

adults, males and females, higher and lower educated individuals, normal controls, and clinical conditions (both neurological and psychiatric). Despite this, other studies revealed that men are more likely to need help in some activities (preparing meals, doing laundry, and taking medications; Niti, Ng, Chiam, & Kua, 2007), although there were some evidence against the DIF effect related to gender in items related to shoulder functional status (Crane, Hart, Gibbons, & Cook, 2006). The DIF effect related to age was also detected in some studies (LaPlante, 2010; Niti et al., 2007). For example, older elderly are more likely to need help in preparing meals (Niti et al., 2007). Additionally, LaPlante (2010) concludes that DIF effects by age are balanced and do not bias the measure.

## Conclusions

In Portugal, the absence of systematic research in adapting and validating instruments for the functional capacity assessment led to the development of the IAFAI. Specifically, only a few functional assessment instruments have some type of validation studies for Portuguese population, for example, the Barthel Index (Araújo, Pais Ribeiro, Oliveira, & Pinto, 2007), and the Lawton & Brody Instrumental Activities of Daily Living Scale (Araújo, Pais Ribeiro, Oliveira, Pinto, & Martins, 2008). The main advantages of the IAFAI were (i) the exam of both BADLs and IADLs, (ii) the content of the items were appropriate to the Portuguese population, and (iii) with several validation and normalization studies to demonstrate its psychometric characteristics. In order to accomplish this, we already performed an initial exploratory study (Sousa et al., in press). In this article, we intended to study the psychometric characteristics of the IAFAI and develop the final version of the inventory regarding their items and response categories. The results of this study suggest that the IAFAI is a comprehensive and useful instrument to assess functional incapacity because it reveals good values of internal consistency, results in adequate person and item separation reliability indexes, and is able to differentiate between normal controls and clinical patients. The consolidation of the original IAFAI categories into dichotomous categories demonstrates a better model fit and increases the reliability indexes. The DIF analysis also demonstrated the IAFAI generalized validity according to important variables (group, age, gender, and education).

Future studies should validate the IAFAI in specific clinical conditions, such as traumatic brain injury, mild cognitive impairment, and Alzheimer's disease as well as establish the normative parameters for the Portuguese population considering important variables such as gender, age, and medical

conditions. Additional studies should consider other psychometric studies regarding rater reliability, test–retest stability, and follow-up studies in some clinical conditions (mild cognitive impairment and dementia) as well as the development of a short form.

## **Appendix A**

### *Adults and Older Adults Functional Assessment Inventory (IAFAI)*

1. Taking meals at appropriate time.
2. Taking a full glass to mouth without spilling.
3. Eating properly without spilling the food.
4. Cutting foods to feed.
5. Dressing.
6. Undress.
7. Buttoning clothes.
8. Tying the shoelaces.
9. Bathing the whole body.
10. Brushing teeth.
11. Washing and brushing hair.
12. Cutting nails.
13. Going to the bathroom when needed.
14. Moving into or near house without losing balance or falling.
15. Up and down stairs.
16. Reaching objects above the shoulder line.
17. Moving in or out of the bed safely.
18. Getting in and out of the bath safely.
19. Dialing a phone number.
20. Answering the phone.
21. Transmit a message.
22. Understanding what people say.
23. Holding a conversation with someone.
24. Having a snack.
25. Cooking a meal.
26. Washing dishes.
27. Doing laundry.
28. Cleaning the floor.
29. Dealing with trash.
30. Caring for plants, yard, or animals.
31. Having contacts for emergencies.

32. Remembering where important objects are (e.g. keys, documents, or money).
33. Turn off the stove, oven, or iron.
34. Check if the water taps are closed.
35. Check if the doors and windows are closed.
36. Ensure that the heater or fireplace pose no danger.
37. Telling someone the main aspects of a news listen on TV.
38. Reading and understanding a book or newspaper.
39. Be careful to pick a recipe or buy medication beforehand.
40. Going to medical appointments and explaining clearly why.
41. Taking medications as prescribed.
42. Meet and deal well with money.
43. Using ATM cards.
44. Completing a check properly.
45. Organize finances to paying bills.
46. Going to the bank to care for finances.
47. Going out without getting lost.
48. Using public transportation when needed.
49. Plan and organize something with family or friends.
50. Continue to perform some usual activities.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Portuguese Science and Technology Foundation through a PhD grant (SFRH/BD/47677/2008) that was awarded to the first author and by the Calouste Gulbenkian Foundation through the project "Validation of memory tests, functional assessment and quality of life inventories" (Process 74569; SDH 22 Neurosciences).

### **References**

- Andrich, D. (1978). A rating formulation for order response categories. *Psychometrika*, 43, 561–573.
- Andrich, D. (1988). *Rasch models for measurement*. London, England: Sage.
- Andrich, D., de Jong, J., & Sheridan, B. E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost & R. Langeheine. (Eds.),

- Applications of latent trait and latent class models in the social sciences* (pp. 59–72). Munster, Germany: Waxmann Verlag.
- Araújo, F., Pais Ribeiro, J. L., Oliveira, A., & Pinto, C. (2007). Validação do Índice de Barthel numa amostra não institucionalizada [Validation of the Barthel Index in a community-dwelling older adults sample]. *Revista Portuguesa de Saúde Pública, 25*, 59–66.
- Araújo, F., Pais Ribeiro, J. L., Oliveira, A., Pinto, C., & Martins, T. (2008). Validação da escala de Lawton e Brody numa amostra de idosos não institucionalizados [Validation of the Lawton and Brody scale in a community-dwelling older adults sample]. *Proceedings of the 7th National Health Psychology Congress, 217–220*.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57*, 289–300.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bourdel-Marchasson, I., Helmer, C., Fagot-Campagna, A., Dehail, P., & Joseph, P. A. (2007). Disability and quality of life in elderly with diabetes. *Diabetes and Metabolism, 33*, S66–S74.
- Breithaupt, K., & McDowell, I. (2001). Considerations for measuring functioning of the elderly: IRM dimensionality and scaling analysis. *Health Services and Outcomes Research Methodology, 2*, 37–50.
- Burns, A., Lawlor, B., & Craig, S. (2004). *Assessment scales in old age Psychiatry* (2nd ed.). London, England: Taylor & Francis Group.
- Crane, P. K., Hart, D. L., Gibbons, L. E., & Cook, K. F. (2006). A 37-item shoulder functional status item pool had negligible differential item functioning. *Journal of Clinical Epidemiology, 59*, 478–484.
- Fieo, R. A., Austin, E. J., Starr, J. M., & Deary, I. J. (2011). Calibrating ADL-IADL scales to improve measurement accuracy and to extend the disability construct into the preclinical range: a systematic review. *BMC Geriatrics, 11*, 42.
- Finlayson, M., Mallinson, T., & Barbosa, V. M. (2005). Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) items were stable over time in a longitudinal study on aging. *Journal of Clinical Epidemiology, 58*, 338–349.
- Galasko, D., Bennett, D., Sano, M., Ernesto, C., Thomas, R., Grundman, M., & Ferris, S. (1997). An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's disease cooperative study. *Alzheimer Disease and Associated Disorders, 11*, 33–39.
- Gélinas, I., Gauthier, L., McIntyre, M., & Gauthier, S. (1999). Development of a functional measure for persons with Alzheimer's disease: The Disability Assessment for Dementia. *American Journal of Occupational Therapy, 53*, 471–481.

- Gill, T. M. (2010). Assessment of function and disability in longitudinal studies. *Journal of the American Geriatrics Society*, 58, 308–312.
- Granger, C. V., Deutsch, A., & Linn, R. T. (1998). Rasch analysis of the Functional Independence Measure (FIM) Mastery Test. *Archives of Physical Medicine and Rehabilitation*, 79, 52–57.
- Green, M. F., Kern, R. S., & Heaton, R. K. (2004). Longitudinal studies of cognition and functional outcome in schizophrenia: Implications for MATRICS. *Schizophrenia Research*, 72, 41–51.
- Hobart, J., & Cano, S. (2009). The Rasch measurement model. *Health Technology Assessment*, 13, 19–32.
- Hoeymans, N., Feskens, E. J. M., van den Bos, G. A. M., & Kromhout, D. (1996). Measuring functional status: Cross-sectional and longitudinal associations between performance and self-report. *Journal of Clinical Epidemiology*, 49, 1103–1110.
- Holland, P., & Thayer, D. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun. (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: LEA.
- Instituto Nacional de Estatística. (2009). *Projeções da população residente em Portugal 2008–2060* [Projections about the Portuguese population 2008–2060]. Lisboa, Portugal: Author.
- Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association*, 185, 914–919.
- Keith, R. A., Granger, C. V., Hamilton, B. B., & Sherwin, F. S. (1987). The functional independence measure: A new tool for rehabilitation. *Advances in Clinical Rehabilitation*, 1, 6–18.
- Knutsson, I., Rystrom, H., Reimer, J., Nyberg, P., & Hagell, P. (2010). Interpretation of response categories in patient-reported rating scales: A controlled study among people with Parkinson's disease. *Health and Quality of Life Outcomes*, 8, 61–69.
- Landi, F., Onder, G., Casari, M., Zamboni, V., Russo, A., & Barillaro, C., ... Silvernet-HC Study Group. (2006). Functional decline in frail community-dwelling stroke patients. *European Journal of Neurology*, 13, 17–23.
- LaPlante, M. P. (2010). The classic measure of disability in Activities of Daily Living is biased by age but an expanded IADL/ADL measure is not. *Journal of Gerontology: Social Sciences*, 65B, 720–732.
- Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: Self-maintaining and Instrumental Activities of Daily Living. *Gerontologist*, 9, 179–186.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3, 85–106.
- Linacre, J. M. (2012). A user's guide to WINSTEPS & MINISTEP. Rasch-Model Computer Programs. Program Manual 3.74.0. Chicago, IL: winsteps.com.



- Lindeboom, R., Vermeulen, M., Holman, R., & de Haan, R. J. (2003). Activities of daily living instruments: Optimizing scales for neurologic assessments. *Neurology*, *60*, 738–742.
- Mahoney, F. I., & Barthel, D. W. (1965). Functional evaluation: The Barthel Index. *Maryland State Medical Journal*, *14*, 61–65.
- Marengoni, A., Aguero-Torres, H., Cossi, S., Ghisla, M. K., Martinis, M. D., Leonardi, R., & Fratiglioni, L. (2004). Poor mental and physical health differentially contributes to disability in hospitalized geriatric patients of different ages. *International Journal of Geriatric Psychiatry*, *19*, 27–34.
- Marson, D., & Hebert, K. R. (2006). Functional assessment. In D. K. Attix & K. A. Welsh-Bohmer. (Eds.), *Geriatric neuropsychology: assessment and intervention* (pp. 158–197). New York, NY: The Guilford Press.
- McGrory, S., Shenkin, S. D., Austin, E. J., & Starr, J. M. (2013). Lawton IADL scale in dementia: Can item response theory make it more informative? *Age and Ageing*, *43*, 491–495.
- Moore, D. J., Palmer, B. W., Patterson, T. L., & Jeste, D. V. (2007). A review of performance-based measures of functional living skills. *Journal of Psychiatric Research*, *41*, 97–118.
- Morton, N. A., Keating, J. L., & Davidson, M. (2008). Rasch analysis of the Barthel Index in the assessment of hospitalized older patients after admission for an acute medical condition. *Archives of Physical Medicine and Rehabilitation*, *89*, 641–647.
- Niti, M., Ng, T. P., Chiam, P. C., & Kua, E. H. (2007). Item response bias was present in instrumental activity of daily living scale in Asian older adults. *Journal of Clinical Epidemiology*, *60*, 366–374.
- Østerås, N., Brage, S., Garratt, A., Benth, J. S., Natvig, B., & Gulbrandsen, P. (2007). Functional ability in a population: Normative survey data and reliability for the ICF based Norwegian Function Assessment Scale. *BMC Public Health*, *7*, 278.
- Palacios-Ceña, D., Jiménez-García, R., Hernández-Barrera, V., Alonso-Blanco, C., Carrasco-Garrido, P., & Fernández-de-las-Peñas, C. (2012). Has the prevalence of disability increased over the past decade (2000–2007) in elderly people? A Spanish population-based survey. *Journal of the American Medical Directors Association*, *13*, 136–142.
- Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Chance, J. M., & Filos, S. (1982). Measurement of functional activities of older adults in the community. *Journal of Gerontology*, *37*, 323–329.
- Potter, G. G., & Attix, D. K. (2006). An integrated model for geriatric neuropsychological assessment. In D. K. Attix & K. A. Welsh-Bohmer. (Eds.), *Geriatric neuropsychology: Assessment and intervention* (pp. 5–26). New York, NY: The Guilford Press.
- Prieto, G., & Delgado, A. R. (2007). Measuring Math Anxiety (in Spanish) with the Rasch Rating Scale Model. *Journal of Applied Measurement*, *8*, 149–160.

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research. (Expanded edition, 1980. Chicago: University of Chicago Press).
- Reckase, M. D. (1979). Unifactor latent trait models applied to multi-factor tests: Results and implications. *Journal of Educational Statistics*, 4, 207–230.
- Royall, D., Lauterbach, E., Kaufer, D., Malloy, P., Coburn, K., & Black, K. (2007). The cognitive correlates of functional status: A review from the Committee on Research of the American Neuropsychiatric Association. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 19, 249–265.
- Sauvaget, C., Yamada, M., Fujiwara, S., Sasaki, H., & Mimori, Y. (2002). Dementia as a predictor of functional disability: A four-year follow-up study. *Gerontology*, 48, 226–233.
- Sousa, L. B., Simões, M. R., Firmino, H., & Peisah, C. (2013). Financial and testamentary capacity evaluations: Procedures and assessment instruments underneath a functional approach. *International Psychogeriatrics*, 14, 1–12.
- Sousa, L. B., Simões, M. R., Pires, L., Vilar, M., & Freitas, S. (2008). *Inventário de Avaliação Funcional de Adultos e Idosos (IAFAI): Manual de administração e cotação* [Adults and Older Adults Functional Assessment Inventory (IAFAI): Administration and scoring manual]. Coimbra, Portugal: Faculty of Psychology and Educational Sciences—University of Coimbra.
- Sousa, L. B., Vilar, M., Pires, L., Freitas, S., & Simões, M. R. (in press). Desenvolvimento de um instrumento de avaliação funcional para a população portuguesa: o Inventário de Avaliação Funcional de Adultos e Idosos (IAFAI) [Development of a new instrument for functional assessment in Portuguese population: The Adults and Older Adults Functional Assessment Inventory (IAFAI)].
- Spector, W. D., & Fleishman, J. A. (1998). Combining activities of daily living and instrumental activities of daily living to measure functional disability. *Journal of Gerontology: Social Sciences*, 53B, S46–S57.
- Tennant, A., & Pallant, J. F. (2006). Unidimensionality matters. *Rasch Measurement Transactions*, 20, 1048–1051.
- Tennant, A., Penta, M., Tesio, L., Grimby, G., Thonnard, J. L., Slade, A., . . . Phillips, S. (2004). Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through Differential Item Functioning within the framework of the Rasch model: The PRO-ESOR project. *Medical Care*, 42, 37–48.
- Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment*, 18, 291–307.
- Thomas, V. C., Rockwood, K., & McDowell, I. (1998). Multidimensionality in instrumental and basic activities of daily living. *Journal of Clinical Epidemiology*, 51, 315–321.

- Wada, T., Ishine, M., Sakagami, T., Kita, T., Okumiya, K., Mizuno, K., . . . Matsubayashi, K. (2005). Depression, activities of daily living, and quality of life of community-dwelling elderly in three Asian countries: Indonesia, Vietnam, and Japan. *Archives of Gerontology and Geriatrics*, *41*, 271–280.
- Walker, J., Böhnke, J. R., Cerny, T., & Strasser, F. (2010). Development of symptom assessments utilizing item response theory and computer-adaptative testing: A practical method based on a systematic review. *Critical Reviews in Oncology/Hematology*, *73*, 47–67.
- West, S. K., Rubin, G. S., Munoz, B., Abraham, D., Fried, L. P., & Salisbury Eye Evaluation Project Team. (1997). Assessing functional status: Correlation between performance on tasks conducted in a clinical setting and performance on the same task conducted at home. *Journal of Gerontology: Medical Sciences*, *52A*, 209–217.
- Wood, K., Edwards, J., Clay, O., Wadley, V., Roenker, D., & Ball, K. (2005). Sensory and cognitive factors influencing functional ability in older adults. *Gerontology*, *51*, 131–141.
- World Health Organization. (2000). *WHODAS-II Disability Assessment Schedule: Training manual, a guide to administration*. Geneva, Switzerland: Author.
- World Health Organization. (2001). *International classification of functioning, disability and health: ICF*. Geneva, Switzerland: Author.
- World Health Organization. (2002). *Active aging: A police framework*. Geneva, Switzerland: Author.
- Wright, B. D. (1999). Model selection: Rating scale or partial credit? *Rasch Measurement Transactions*, *12*, 641–642.
- Wright, B. D., & Douglas, G. A. (1976). *Rasch item analysis by hand* (Research Memorandum No 21). Statistical Laboratory, Department of Education, University of Chicago, Chicago, IL.
- Yeh, Y. C., Lin, K. N., Chen, W. T., Lin, C. Y., Chen, T. B., & Wang, P. N. (2011). Functional disability profiles in amnesic Mild Cognitive Impairment. *Dementia and Geriatric Cognitive Disorders*, *31*, 225–232.
- Zwick, R., & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement*, *26*, 55–66.

## Author Biographies

**Liliana B. Sousa** holds master's degree in psychology (psychological assessment, counselling, and rehabilitation) and is currently a PhD student (Forensic Psychology), Faculty of Psychology and Educational Sciences at the University of Coimbra. She received her PhD grant from the Science and Technology Foundation (SFRH/BD/

47677/2008) for the project “Functional approach in determining the financial and testamentary capacity: guidelines and development of assessment instruments.”

**Gerardo Prieto** is a full professor, the Faculty of Psychology, at the University of Salamanca, Spain (psychometric methods). His research is mainly related to the application of psychometric models for the development of tests (academic achievement, psychological, and neuropsychological) and visuospatial abilities (predictive validity, sex differences, and training).

**Manuela Vilar** holds a master’s degree in psychology (pedagogic psychology), and is currently an assistant professor, Faculty of Psychology and Educational Sciences at the University of Coimbra, Portugal (psychological assessment in forensic and gerontology contexts). Vilar is a member of the Portuguese WHOQOL team—adaptation, validation, and normalization of the WHOQOL-OLD for the Portuguese elderly population.

**Horácio Firmino** is a psychiatrist and clinical head. He has been involved in the development of psychogeriatrics for over 10 years through research, education, and training as well as clinical care. He was one of the founders of the Portuguese Psychogeriatric Association, joined the International Psychogeriatrics Association (IPA) in 1993, and has participated in many of its activities.

**Mário R. Simões** is a full professor, Faculty of Psychology and Educational Sciences at the University of Coimbra, Portugal (neuropsychological assessment in forensic and gerontology contexts). His research is mainly related to the adaptation and validation of neuropsychological assessment instruments in forensic and gerontology contexts.