CULTURAL DIVERSITY AND THE MEASUREMENT OF FUNCTIONAL IMPAIRMENT: A CROSS-CULTURAL VALIDATION OF THE AMSTERDAM IADL QUESTIONNAIRE

Mark A. Dubbelman a*, Merike Verrijp a, David Facal b, Gonzalo Sánchez-Benavides c, Laura J.E. Brown d, Wiesje M. van der Flier a,e, Hanna Jokinen f, Athene Lee g, Iracema Leroi d, Cristina Lojo-Seoane b, Vuk Milošević h, José Luis Molinuevo i, Arturo X. Pereiro Rozas b, Craig Ritchie i, Stephen Salloway f, Gemma Stringer d, Stelios Zygouris j,k, Bruno Dubois j, Stéphane Epelbaum l, Philip Scheltens a, Sietske A. M. Sikkes a,e

Affiliations
a Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands
b Department of Developmental Psychology, University of Santiago de Compostela, A Coruña, Spain
c Barcelonaβeta Brain Research Center (BBRC), Pasqual Maragall Foundation, Barcelona, Spain
d Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Science Centre, Manchester, United Kingdom
e Department of Epidemiology and Biostatistics, Amsterdam UMC, Amsterdam, The Netherlands
f Clinical Neurosciences, Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
g Butler Hospital, Warren Alpert Medical School of Brown University, Providence, RI, United States of America
h Clinic of Neurology, Clinical Center Niš, Niš, Serbia
i University of Edinburgh, Edinburgh, United Kingdom
j School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece
k Network Aging Research, Heidelberg University, Heidelberg, Germany
l Département de neurologie, Institut de la Mémoire et de la Maladie d’Alzheimer (IM2A) of the Pitié-Salpêtrière Hospital & ARAMIS, Sorbonne Université, Inria de Paris, Institut du cerveau et de la moelle épinière (ICM), Paris, France

Corresponding author
M. A. Dubbelman, MSc
Amsterdam UMC
Alzheimer Center, Department of Neurology
P.O. Box 7057
1007 MB Amsterdam
The Netherlands
Phone: +31 20 4448527
E-mail: m.dubbelman@amsterdamumc.nl

Keywords: instrumental activities of daily living; dementia; Alzheimer’s disease; functional decline; cross-cultural validation; diversity; item response theory; differential item functioning
Abstract

INTRODUCTION: Assessment of cognitively complex instrumental activities of daily living (IADL) is important in (incipient) dementia, but little is known about the influence of cultural differences. We aim to investigate the cross-cultural comparability of IADL.

METHODS: IADL were measured in 3,571 individuals (67.1 ± 9.5 years old, 44.7% female) in 11 international studies using the Amsterdam IADL Questionnaire (A-IADL-Q). Cultural, age and gender bias were assessed on item level using differential item functioning (DIF) and on scale level by calculating correlations between IADL and cognitive and functional measures.

RESULTS: We did not find meaningful bias on item or on scale level, as indicated by low DIF effect sizes ($R^2$ range 0–0.03) and correlations in expected directions and of similar magnitude.

DISCUSSION: Our results suggest that IADL may be universally suitable as outcome measure. The A-IADL-Q seems robust to demographic diversity and is capable of reliably measuring IADL impairment.
1. Introduction

Impairment in cognitively complex everyday activities, such as doing grocery shopping, managing personal finances and using mobile devices, are among the first symptoms of dementia [1-3]. These activities are referred to as ‘instrumental activities of daily living’ (IADL). IADL performance is related to quality of life, caregiver burden and resource utilization [4]. Moreover, IADL impairment in preclinical stages might be a predictor of progression to dementia [5]. Functional impairment can be an important outcome measure for clinical trials and clinical practice, and in recently drafted industry guidelines, the U.S. Food and Drug Administration recommended the use of functional impairment as a measure for effectiveness of treatment and of disease progression [6]. When using IADL as a functional outcome measure across culturally diverse populations, one must consider the fact that everyday activities differ between countries and cultures. Nevertheless, little is known about the cross-cultural comparability of IADL and its measures.

Scientific literature concerning diversity and resulting cultural and ethnoracial disparities in the context of AD is scarce [7, 8]. Culture may have a diverse impact on cognition and, in extension, on functional status [8]. Mere translation of an instrument does not always account for these disparities [9, 10], and while many functional instruments have been translated into numerous languages, there is no gold standard for cross-cultural adaptation of questionnaires [11]. While a few studies have highlighted the potential of IADL as a global outcome measure in dementia research [12, 13], frequently used IADL instruments often include potentially culture-specific activities, such as balancing a checkbook [14-16]. In addition, several instrument-specific cross-cultural adaptation studies have reported potential bias due to differences in the distribution of IADL between the genders: traditional instruments, such as the Lawton & Brody scale [2], may rely too heavily on household tasks, which are predominantly performed by women [13, 17-19]. This emphasizes the importance of determining whether all activities are cross-culturally relevant.
A modern instrument for measuring IADL is the Amsterdam IADL Questionnaire (A-IADL-Q). It was developed to capture the earliest changes in IADL in (preclinical) dementia. Previous studies have shown that the A-IADL-Q has good internal consistency [20], is independent of age and gender [21], and has good validity and reliability [20-22] and sensitivity to change over time [23]. A short version of the questionnaire (A-IADL-Q-SV) was recently developed to create a more concise measure, and also aimed to reduce potential cultural bias by only including widely relevant activities [24]. International use of the A-IADL-Q is steadily increasing. All translations have gone through a cross-cultural adaptation process in which experts and prospective users were asked to evaluate the translated instrument.

We aimed to investigate the comparability of IADL across culturally diverse groups by studying item bias in the A-IADL-Q between eight Western countries: The Netherlands, Spain, France, United States, United Kingdom, Greece, Serbia, and Finland. As an additional step, we investigated the relations between IADL and cognitive and functional measures, age, gender, and education.
2. Methods

The present study included data from 3,571 individuals with a completed A-IADL-Q from memory clinics and cognition studies from eight countries: the Netherlands (Amsterdam Dementia Cohort [25] and European Prevention of Alzheimer’s Dementia Longitudinal Cohort Study, EPAD [26, 27]), Spain (Compostela Aging Study [28, 29]; EPAD; and ALFA project [30]), France (INSIGHT pre-AD [31]; EPAD; and Socrates study), United States (Butler Alzheimer’s Prevention Registry [32]), United Kingdom (EPAD and SAMS project [33]), Greece (Greek Association for Alzheimer’s Disease and Related Disorders), Serbia (Niš Clinic of Neurology [34]), and Finland (Helsinki Small Vessel Disease study).

Participants had some degree of cognitive complaints, or had an increased genetic or neurovascular risk for cognitive decline. Participants were recruited from memory clinics, through advertisement, or from existing databanks. Inclusion criteria ranged from being cognitively normal to having a dementia-related diagnosis. Other relevant inclusion and exclusion criteria for each cohort in this study can be found in Table 1. Participants provided written informed consent, and the studies were approved by their institutional review boards, which included, in each, consent for data sharing.
Table 1: Information about participants, in- and exclusion criteria, and information about the A-IADL-Q administration per included sample

| Study name | Country | Age range included | Research environment | Participants included | Recruitment | Relevant inclusion and exclusion criteria | A-IADL-Q version | Clinical measures |
|------------|---------|--------------------|----------------------|-----------------------|-------------|------------------------------------------|-----------------|------------------|
| Amsterdam Dementia Cohort [25] | Netherlands | 25–84 years | Consecutive memory clinic patients | 1,429 | Original (n = 730) | Cognitive complaints without dementia; Age ≥ 50 years | Original (n = 730) | MMSE, CAMCOG, CDR, GDS |
| Compostela Aging Study [28, 29] | Spain | 50–101 years | MCI patients referred by GP | 600 | Original | No dementia; Age ≥ 50 years | Original | MMSE, CAMCOG, CDR, GDS |
| European Prevention of Alzheimer’s Dementia Longitudinal Cohort Study (EPAD) [26, 27] | Spain | 51–88 years | Participants from existing study cohorts | 480 | Original | CN (MMSE ≥ 26, CDR 0); No neurological diseases; Age 45–74 years | Original | MMSE, CAMCOG, CDR, GDS |
| ALFA+ Study [30] | France | 70–85 years | Mostly offspring of AD patients | 333 | SV | CN (MMSE ≥ 27, CDR 0); Amyloid PET at baseline; No episodic memory deficits, no neurological diseases; Not living in nursing home; Age 70–85 years | SV | MMSE, CAMCOG, CDR, GDS |
| INSIGHT preAD [31] | France | 70–85 years | Advertisement recruited | 308 | SV | CN or mild memory loss; No neurological diseases or dementia diagnosis; Age 55–85 years | SV | MMSE, CAMCOG, CDR, GDS |
| Butler Alzheimer’s Prevention Registry [32] | United States of America | 58–77 years | Memory clinic patients | 154 | SV | Dementia-related diagnosis (MMSE ≥ 10); No neurological diseases other than dementia; Age 40–85 years | SV | MMSE, GDS |
| SOCRATES | France | 64–85 years | Patients from day center for dementia | 98 | SV | Dementia-related diagnosis; Reliable informant; No neurological diseases other than dementia; Age ≥ 65 years | SV | MMSE, GDS |
| Greek Association of Alzheimer’s Disease and Related Disorders | Greece | 65–92 years | Memory clinic patients | 61 | SV | CN, MCI, post-stroke cognitive impairment | SV | MMSE, GDS |
| Nií Clinic of Neurology [34] | Serbia | 26–93 years | Patients with neuroimaging data selected from existing databank | 45 | SV | No major neurological symptoms or psychiatric disease; Independence in basic ADL; No large infarcts, hemorrhages, contusion or tumor on MRI; Age 65–75 years | SV | MMSE, GDS |
| Helsinki Small Vessel Disease study | Finland | 66–75 years | Recruited from dementia research registry, memory clinic patients | 43 | SV | SCD (ECoG ≥ 1.436 and answered “yes” when asked if “concerned they have a memory or other thinking problem”), MCI; Age ≥ 65 years | SV | MMSE, GDS |
| SAMS Project [33] | United Kingdom | 65–82 years | | 22 | | | | |

* Participants living in a nursing home were excluded from validation and no clinical measures were obtained for them. Additional inclusion and exclusion criteria are available upon request.

Abbreviations: MCI mild cognitive impairment; GP general practitioner; AD Alzheimer’s disease; CN cognitively normal; MMSE Mini-Mental State Examination; CDR Clinical Dementia Rating; ADL activities of daily living; MRI magnetic resonance imaging; SCD subjective cognitive decline; CAMCOG Cambridge Cognitive Examinations; GDS Geriatric Depression Scale
2.1 Measures

2.1.1 Amsterdam IADL Questionnaire (A-IADL-Q)

The Amsterdam IADL Questionnaire (A-IADL-Q) assesses cognitively complex IADL that are prone to decline in incipient dementia and covers a wide range of activities. The original version consists of 70 items, while the short version has 30. The studies included here used the original ($n = 6$) or the short version ($n = 7$). We analyzed both versions, with a special focus on the short version (A-IADL-Q-SV), because all items from the short version are also included in the original, and can therefore be compared between all participants.

The questionnaire is scored using item response theory (IRT), as described elsewhere [20, 22]. IRT scores are considered more precise and are less affected by missing data, compared to sum scores [35]. All items have five response categories, ranging from having ‘no difficulty’ in performing an activity to being ‘unable to perform’ an activity due to cognitive problems. IRT-based $T$-scores were calibrated in a memory-clinic population and were centered around a mean of 50 with a standard deviation (SD) of 10. Lower scores indicate more severe functional impairments.

2.1.2. Clinical Measures

For the construct validation, we excluded individuals living in nursing homes ($n = 130$) because they have limited IADL independence. Mini-Mental State Examination (MMSE, scores range 0–30) [36] and Cambridge Cognition Examination (CAMCOG, scores range 0–107) [37] served as general indications of cognitive functioning. For both measures, lower scores indicate worse cognition. The Clinical Dementia Rating (CDR) [38] was an indicator of functional status. A global CDR score of 0 represents no dementia, and scores of 0.5 to 3 are related to more advanced stages of dementia (and thus more functional impairment). Lastly, the short form Geriatric Depression Scale (GDS, scores range 0–15) [39] was used to assess depressive symptoms, where higher scores are indicative of more severe depressive symptoms. Data was not obtained for all included participants.
2.2 Statistical Analyses
For all analyses, studies were grouped by country, resulting in eight groups. Differences between countries in sociodemographic variables and clinical characteristics were calculated using one-way analysis of variance (ANOVA) or Kruskal-Wallis tests. Post-hoc corrections were performed where applicable.

Confirmatory factor analyses (CFA) were performed on the short version to confirm unidimensionality, which was shown to be applicable to both the Dutch original and the short version of the A-IADL-Q in earlier publications [20, 24]. Structural equation modeling was used with a weighted least squares estimation method [40]. Good model fit was determined based on the following criteria: (a) root mean square error of approximations (RMSEA) ≤ .05 and (b) comparative fit index (CFI) ≥ .90 [41].

We subsequently assessed measurement invariance by investigating item bias using ‘differential item functioning’ (DIF). DIF analysis is a technique for identifying items that have different levels of difficulty and/or discriminatory ability in multiple groups. DIF is assumed to occur when the relationship between a test item and the latent trait (represented in the T-score) is not the same across study-irrelevant groups [35]. It is considered a variation in measurement and is therefore undesirable [42].

We studied DIF between countries, using the Dutch cohort as a reference group for all comparisons. A minimum count of one case in at least two different response categories was required in each country for every item. We used DIF analyses based on ordinal logistic regressions, in which a null model and three hierarchically nested models are created and compared for each item. The logistic regression procedure was chosen because it is more accurate than other DIF detection procedures when the reference group is large (e.g., \( n = 1,500 \)), even when focal groups are smaller (e.g., \( n = 100 \)) [43]. When DIF is present and constant across all levels of the latent trait, it is called uniform DIF. An item with uniform DIF may be easier or more difficult in one group than in the other [44]. When an item is easier at one level of the latent trait and more difficult at another level, it is considered to have non-uniform DIF [44]. We used a two-pronged detection criterion for establishing presence of significant and
meaningful DIF [45]: (1) a statistically significant likelihood-ratio chi-square test with an α level of .01, allowing for a 1% false positive identification rate, and (2) a corresponding McFadden’s pseudo $R^2$ value of .035 or larger. Furthermore, we used the following effect size criteria to quantify DIF size: $R^2$ values between .035 and .070 for moderate, and above .070 for large DIF [46].

In addition to investigating DIF between countries, we also analyzed DIF between men and women, young and old participants, and high and low education in the entire sample. The division between young and old age and high and low education was made on the basis of median split, with participants under the median age being classified as young, and under the median years of education as having low education. We used the ‘lordif’ package for R, developed by Choi, Gibbons [42].

As a means of construct validation, Pearson’s $r$ for continuous or Kendall’s $\tau$ correlation coefficients for ordinal-level measures were calculated for the association between A-IADL-Q-SV T-scores and age, education level, gender of the participant, cognitive functioning (MMSE and CAMCOG), functional state (CDR), and mood (GDS).

Significance levels were set at $p < .05$, unless otherwise indicated. Data were processed in SPSS Statistics version 22 [47], R version 3.5.2 [48], and Mplus 7 [40].
3. Results

On average, participants were 67.1 ± 9.5 (m ± SD) years old. The mean age differed between countries ($F(7, 3420) = 93.22, p < .001$): it was the highest in Greece (80.0 ± 6.4 years old) and the lowest in the Netherlands (63.8 ± 8.5 years old). The distribution of men and women differed between countries ($\chi^2(7) = 124.56, p < .001$): the proportion of women was highest in the United States (67.5%) and lowest in Greece (29.5%). Participants from each country also differed from each other in years of education ($F(7, 2969) = 73.68, p < .001$). Participants in the U.S. had the most years of education (16.8 ± 2.3 years), whereas participants in Greece had the least (9.5 ± 4.3 years). Table 2 shows the demographics and clinical measures per country.

Table 2

| Demographic and clinical characteristics and differences between countries |
|-------------------------------------------------------------|
| All | Netherlands | Spain | France | USA | UK | Greece | Serbia | Finland |
|---|---|---|---|---|---|---|---|---|
| Total n | 3,571 | 1,515 | 1,151 | 509 | 154 | 93 | 61 | 45 | 43 |
| Females, n (%) | 1,597 (44.7) | 637 (42.0) | 485 (51.5) | 262 (67.5) | 104 (46.2) | 18 (29.5) | 25 (55.6) | 23 (53.5) | $p < .001$ |
| Age, years | 67.14 ± 9.5 | 63.78 ± 8.5 | 67.84 ± 10.4 | 73.48 ± 6.2 | 68.42 ± 4.5 | 79.99 ± 6.4 | 65.44 ± 4.5 | 71.69 ± 2.8 | $p < .001$ |
| Dementia diagnosis, n (%) | 1860 (29.9) | 647 (47.2) | 188 (20.2) | 0 (0) | 0 (0) | 21 (80.8) | 4 (8.9) | 0 (0) | $p < .001$ |
| A-IADL-Q T-score | 58.40 ± 14.2 | 51.54 ± 11.7 | 61.82 ± 15.2 | 67.33 ± 9.4 | 67.48 ± 3.5 | 71.16 ± 5.1 | 39.48 ± 13.9 | 61.67 ± 8.8 | 66.30 ± 5.2 | $p < .001$ |
| Clinical measures | | | | | | | | | |
| MMSE | 26.20 ± 4.6 | 24.22 ± 5.0 | 27.76 ± 3.7 | 28.62 ± 1.2 | 29.35 ± 1.0 | 28.46 ± 1.5 | 19.58 ± 4.6 | 27.49 ± 3.6 | 27.60 ± 2.2 | $p < .001$ |
| CAMCOG | 78.57 ± 17.3 | 78.75 ± 16.1 | 80.98 ± 19.1 | — | — | 41.62 ± 9.7 | — | — | $p < .001$ |
| CDR, M (IQR) | 0 (0–0.5) | 0.5 (0–1) | 0 (0–0) | 0 (0–0) | 0 (0–2) | 2 (0.5–2) | — | — | $p < .001$ |
| GDS | 3.66 ± 3.6 | 3.80 ± 3.3 | 4.09 ± 4.0 | 4.33 ± 4.2 | 0.85 ± 1.3 | 3.52 ± 4.5 | 2.38 ± 3.1 | — | 2.10 ± 3.1 | $p < .001$ |
All data are displayed as mean ± standard deviation, except as stated otherwise. Percentages shown are valid percentages (excluding missing values). Differences between groups were calculated using one-way analysis of variance, Pearson chi-square or Kruskal-Wallis test, as appropriate. "—" denotes that data was not available.

1 Data was not obtained for all participants.

2 The score shown is the average of the scores for the version (original or SV) that was administered.

Abbreviations: M median; IQR interquartile range; A-IADL-Q Amsterdam IADL Questionnaire; MMSE Mini-Mental State Examination; CAMCOG Cambridge Cognitive Examinations; CDR Clinical Dementia Rating; GDS Geriatric Depression Scale

The overall mean score on the A-IADL-Q was 58.40 ± 14.2. The scores were different between countries \((F(7, 3477) = 198.07, p < .001)\) and were highest in the United Kingdom \((71.16 \pm 5.1)\) and lowest in Greece \((39.48 \pm 13.9)\). A-IADL-Q scores per country are shown in Table 2.

### 3.1 Endorsement

![Endorsement Chart](image)

**Figure 1:** Overall, item endorsement was similar between countries, except for the three items displayed here: 'Using a computer', 'Driving a car', and 'Using public transportation'. The percentage of participants in each country that performed these activities in the four weeks preceding completion of the A-IADL-Q is displayed here, sorted by the level of endorsement, which ranged from 99.4% to 8.2% across all three items.

Generally, item endorsement was comparable between countries, except for 3 items. Participants from Greece (8.2%), Spain (52.7%), and Serbia (53.3%) endorsed ‘using a computer’ less than those from other countries (range 75.2%–97.4%). ‘Using public transportation’ was less frequently endorsed by participants from the United States (27.9%), whereas relatively more Americans had driven a car (99.4%) than individuals in the other countries (see Figure 1).
CFA in the entire sample showed a good model fit for a single factor in the short version of the A-IADL-Q (RMSEA = .019, 95%CI = [.018, .021]; CFI = .999). A good model fit was also found in the cohorts from Spain (RMSEA = .058, 95%CI = [.055, .060]; CFI = .997), and France (RMSEA = .023, 95%CI = [.017, .029]; CFI = .998). Due to lack of variance in item responses in multiple items in all other countries, CFA could not be performed reliably using all items of the A-IADL-Q-SV, and is therefore not reported here.

### 3.2 Item bias

![Diagram of item bias in A-IADL-Q-SV](image)

**Figure 2: Item bias in the A-IADL-Q-SV.**

Mapping of DIF effect sizes from the A-IADL-Q-SV. Each country was compared to the Netherlands. Negligible DIF is displayed in green, meaningful DIF with a large effect is shown in red (but was not present). Items that could not be analyzed (due to limited response variation) are displayed in gray.

Regarding item bias, no meaningful DIF was found in the A-IADL-Q-SV (all $R^2 < .035$, range .000–.034). Figure 2 shows the $R^2$ effect sizes of DIF for all analyzed items. In the original version, four items had meaningful DIF with a moderate effect. Three of the four items had uniform DIF, and they were found
in the Spanish group: ‘using the washing machine’ ($R^2 = 0.043$), ‘making appointments’ ($R^2 = 0.064$), and ‘playing card and board games’ ($R^2 = 0.043$). The first item was more difficult for Spanish individuals, the other two were easier, as compared to the Dutch reference group. The fourth item had non-uniform DIF and was found in the French group: ‘functioning adequately at work’ ($R^2 = 0.064$). The item appeared to be better at discriminating between people with lower and higher levels of functional impairment in France than in the Netherlands. We used the DIF results to re-estimate the $T$-scores for Spanish and French participants, thus correcting for the effect of DIF. In the Spanish group, the mean score decreased by 0.16 points (changes ranged $-1.14 - +0.43$); in the French group, the mean score decreased by 0.07 points (changes ranged $-1.33 - +0.35$). These changes corresponded to a difference of less than 1/10 of a SD, and can therefore be considered negligible.

In the entire sample, neither the original nor the short version showed DIF for age, gender, or education.

### 3.3 A-IADL-Q-SV construct validation

Overall, all correlations were in the expected directions and of similar magnitudes as compared to the original validation data from the Netherlands. Age seemed more strongly associated with IADL impairment than expected in Spain ($r = -0.47$, 95%CI = [-0.51, -0.42]), Greece ($r = -0.31$, 95%CI = [-0.52, -0.06]), and Serbia ($r = -0.48$, 95%CI = [-0.68, -0.21]) than in the Netherlands ($r = -0.08$, 95%CI = [-0.13, -0.02]). MMSE scores appeared to be less associated with IADL impairment France ($r = 0.11$, 95%CI = [0.02, 0.21]), United States ($r = 0.12$, 95%CI = [-0.05, 0.27]), and United Kingdom ($r = -0.10$, 95%CI = [-0.33, 0.14]), compared to the reference ($r = 0.33$, 95%CI = [0.28, 0.38]). In these countries, the MMSE had a restricted score range. Conversely, MMSE scores were more strongly associated with IADL impairment in Serbia ($r = 0.56$, 95%CI = [0.32, 0.73]). An overview of all correlations can be found in the Appendix.
4. Discussion

In this study, we demonstrated that the influence of cultural diversity on the measurement of IADL impairment, as measured with the A-IADL-Q, is minimal. We found limited differences between countries in terms of activity endorsement. Moreover, there was no evidence of item bias for country, suggesting that the A-IADL-Q consistently measures IADL impairment. Furthermore, we found no item bias for age, gender, or education. Associations with demographic, cognitive and functional measures were also in the expected directions and largely in line with earlier hypothesized magnitudes.

Diversity research in the context of IADL and dementia is very scarce [8]. A few studies have shown a general comparability of IADLs across cultural and ethnoracial diversity [12, 13], while others have shown differences between cultures, genders, and age [49, 50], albeit not in dementia-specific contexts. For optimal comparison of functional outcome in international studies and clinical trials, a reliable, cross-culturally validated instrument is crucial.

Using the A-IADL-Q-SV, we did not find meaningful DIF, indicating there is no evidence of item bias for cultural diversity. This suggests that the A-IADL-Q yields reliable and cross-nationally comparable estimations of functional decline. This absence of item bias could be the result of the development of the A-IADL-Q-SV, in which international experts provided feedback on the cross-cultural comparability of the items [24]. This is also reflected by the generally similar activity endorsement across countries. Only minor differences in endorsement were found in the A-IADL-Q-SV activities. Furthermore, the translations of the questionnaires were subject to a cross-cultural adaptation process (such as described for the Spanish translation in Facal, Carabias [28]). Because the A-IADL-Q-SV does not appear to have item bias, scores do not need to be adjusted in order to be compared across countries. In the original version, a few items appeared to be biased in Spain and France. ‘Making appointments’ had the largest DIF effect, and a potential explanation for this is that examples were added in the Spanish translation, because language experts indicated that the proposed translation for the word
'appointments' (citas) could be interpreted as ‘(romantic) dates’. The other items with DIF had a smaller effect, and no clear reason for the presence of DIF could be discerned. Despite the finding of item bias in the original version, the effect on the total scores was minimal. We therefore consider the bias negligible, evidenced as well by the absence of DIF in the short version. The A-IADL-Q-SV might be the preferred version for future international use, as it includes only the most broadly relevant everyday activities, does not have item bias, has comparable construct validity, and is more pragmatic.

Moreover, there was no item bias for age, gender, or education. This further underlines the independence of the estimations of functional disability from these confounders, confirming previous studies on the A-IADL-Q [21, 22, 24]. This is an important finding, because other functional instruments do appear to be biased for confounders [49, 50].

As an additional step, we performed a construct validation of the A-IADL-Q-SV. Overall, associations with age, gender, education and mood were small, while we found moderate correlations with cognition and functional state. These findings correspond to the associations previously described for the original version [21]. The variation in associations between countries may be study-specific: demographic, cognitive and functional characteristics of participants were different in each country. In Spain, Greece, and Serbia, participants were older than average, and associations between age and IADL were stronger. In Spain, an association between age and IADL functioning was found earlier in a group of patients without dementia [28]. In France, the United States and the United Kingdom, the studies recruited mainly cognitively healthy participants, resulting in very limited variation in the measure of cognition and IADL functioning seemed to be less associated with cognitive measures.

An important strength of this study is that we used a data-driven approach to investigate the cross-cultural comparability of IADL. We used DIF, which is a powerful procedure to detect variance in measurement between groups on an item level and was possible as a result of the IRT scoring method.
Not only does DIF tell us whether an item may be biased, but it also provides insight in the impact of the bias on the overall scores and it allows for correction. These advantages allowed us to create a clear picture of possible measurement variance and impact on the instrument. Another strength of the study is that we included data from more than 3,500 individuals from eleven different studies representing eight countries. People with a wide variety of cognitive impairment-related diagnoses or complaints were included, ranging from subjective cognitive decline to dementia. Furthermore, the age of participants ranged from adulthood to old age. The large sample size and large variety in diagnoses and age contributes to the generalizability of our results and conclusions.

This study also had a few limitations. First, we only included data from developed, Western countries, without taking race and ethnicity into account. The present study is an important first step in recognizing the influence of diversity on the measurement of functional impairment, and future studies should build on these findings. Furthermore, it should be noted that ‘culture’ is not interchangeable with ‘country’, because residents of a single country may have different cultural and ethnoracial backgrounds. We operationalized our study by investigating the differences between various cross-culturally adapted translations of the A-IADL-Q. Assuming that the included samples are good random samples of the respective countries’ populations, cultural variation within countries may still be adequately represented by the included cohorts. Second, the variation between the included samples complicated the interpretation of the results. Due to variability in age, sex, and level of cognitive and functional impairment, it is difficult to distinguish group differences caused by culture from those caused by other participant characteristics. We used additional DIF analyses to rule out item bias caused by age, sex, or education, which have been shown to cause bias in other IADL instruments. We did not find such bias, nor did we find cultural bias, indicating that the A-IADL-Q is robust to demographic variability.
To conclude, we validated the A-IADL-Q(-SV) and demonstrated that it is not influenced by culture, age, gender, or education. This is important, because it further underlines the potential of the A-IADL-Q, and the short version in particular, as an outcome measure of daily functioning in clinical practice and clinical trials around the world. Taken together, these results suggest that these IADL may be universal and that the A-IADL-Q is robust to demographic diversity and is capable of reliably measuring IADL impairment.
Acknowledgements
The Amsterdam IADL Questionnaire is free for use in all public health and not-for-profit agencies and can be obtained via https://www.alzheimercentrum.nl/professionals/amsterdam-iadl.

The development of the Amsterdam IADL Questionnaire is supported by grants from Stichting VUmc Fonds and Innovatiefonds Zorgverzekeraars. The Amsterdam Alzheimer Center is supported by Stichting Alzheimer Nederland and Stichting VUmc Fonds. The clinical database structure for the Amsterdam Dementia Cohort was developed with funding from Stichting Dioraphte. The ALFA+ Study has received funding from “la Caixa” Foundation (LCF/PR/GN17/1030004), the Alzheimer’s Association, and an international anonymous charity foundation through the TriBEKa Imaging Platform project. The work for EPAD has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking EPAD grant agreement n° 115736. The authors report no conflict of interests.
Highlights

- Cultural diversity may influence measurement if not adequately accounted for.

- 3,571 people from 8 countries answered the Amsterdam IADL Questionnaire (A-IADL-Q).

- At the item level, only very minor differences were found between eight countries.

- No meaningful bias existed at the scale level; total scores were not influenced.

- The A-IADL-Q can reliably measure IADL impairment across culturally diverse groups.
References

[1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. Arlington, VA: American Psychiatric Association; 2013.
[2] Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. 1969;9:179-86.
[3] Peres K, Helmer C, Amieva H, Orgogozo JM, Rouch I, Dartigues JF, et al. Natural history of decline in instrumental activities of daily living performance over the 10 years preceding the clinical diagnosis of dementia: a prospective population-based study. J Am Geriatr Soc. 2008;56:37-44.
[4] Giebel CM, Sutcliffe C, Challis D. Activities of daily living and quality of life across different stages of dementia: a UK study. Aging Ment Health. 2015;19:63-71.
[5] Luck T, Riedel-Heller SG, Luppa M, Wiese B, Bachmann C, Jessen F, et al. A hierarchy of predictors for dementia-free survival in old-age: results of the AgeCoDe study. Acta Psychiatr Scand. 2014;129:63-72.
[6] Food and Drug Administration. Draft guidance for industry: Alzheimer's disease: developing drugs for the treatment of early stage disease. Silver Spring, MD: Center for Drug Evaluation and Research; 2018.
[7] Chin AL, Negash S, Hamilton R. Diversity and disparity in dementia: the impact of ethnoracial differences in Alzheimer disease. Alzheimer Dis Assoc Disord. 2011;25:187-95.
[8] Babulal GM, Quiroz YT, Albensi BC, Arenaza-Urquijo E, Astell AJ, Babiloni C, et al. Perspectives on ethnic and racial disparities in Alzheimer’s disease and related dementias: Update and areas of immediate need. Alzheimers Dement. 2018.
[9] Parra MA. Overcoming barriers in cognitive assessment of Alzheimer’s disease. Dement Neuropsychol. 2014;8:95-8.
[10] Mungas D, Reed BR, Farias ST, Decarli C. Age and education effects on relationships of cognitive test scores with brain structure in demographically diverse older persons. Psychol Aging. 2009;24:116-28.
[11] Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. J Clin Epidemiol. 2015;68:435-41.
[12] Truscott DJ. Cross-cultural ranking of IADL skills. Ethn Health. 2000;5:67-78.
[13] Nikula S, Jylhä M, Bardage C, Deeg DJH, Gindin J, Minicuci N, et al. Are IADLs comparable across countries? Sociodemographic associates of harmonized IADL measures. Aging Clin Exp Res. 2003;15:451-9.
[14] Galasko D, Bennett D, Sano M, Ernesto C, Thomas R, Grundman M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer’s disease. The Alzheimer’s Disease Cooperative Study. Alzheimer Dis Assoc Disord. 1997;11 Suppl 2:S33-9.
[15] Farias ST, Mungas D, Harvey DJ, Simmons A, Reed BR, Decarli C. The measurement of Everyday Cognition (ECog): Development and validation of a short form. Alzheimers Dement. 2011;7:593-601.
[16] Pfeffer RI, Kurosaki TT, Harrah CH, Jr., Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol. 1982;37:323-9.
[17] Vergara I, Bilbao A, Orive M, Garcia-Gutierrez S, Navarro G, Quintana JM. Validation of the Spanish version of the Lawton IADL Scale for its application in elderly people. Health Qual Life Outcomes. 2012;10:130.
[18] Lechowski L, de Stampa M, Denis B, Tortrat D, Chassagne P, Robert P, et al. Patterns of loss of abilities in instrumental activities of daily living in Alzheimer’s disease: the REAL cohort study. Dement Geriatr Cogn Disord. 2008;25:46-53.
[19] Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? Neurodegener Dis Manag. 2014;4:393-400.
[20] Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, Gillissen F, Romkes R, Knol DL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. 2012;8:536-43.
[21] Sikkes SAM, Knol DL, Pijnenburg YAL, de Lange-de Klerk ES, Uitdehaag BM, Scheltens P. Validation of the Amsterdam IADL Questionnaire(c), a new tool to measure instrumental activities of daily living in dementia. Neuroepidemiology. 2013;41:35-41.
[22] Sikkes SAM, Pijnenburg YAL, Knol DL, de Lange-de Klerk ES, Scheltens P, Uitdehaag BM. Assessment of instrumental activities of daily living in dementia: diagnostic value of the Amsterdam Instrumental Activities of Daily Living Questionnaire. J Geriatr Psychiatry Neurol. 2013;26:244-50.
[23] Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement. 2015;11:1231-40.
[24] Jutten RJ, Peeters CFW, Leijdesdorff SMJ, Visser PJ, Maier AB, Terwee CB, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35.
[25] van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. J Alzheimers Dis. 2018;62:1091-111.
[26] Ritchie CW, Molinuevo JL, Truyen L, Satlin A, Van der Geyten S, Lovestone S. Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. Lancet Psychiatry. 2016;3:179-86.
[27] Solomon A, Kivipelto M, Molinuevo JL, Tom B, Ritchie CW, EPAD Consortium. European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol. BMJ Open. 2018;8.
[28] Facal D, Carabias MAR, Pereiro AX, Lojo-Seoane C, Campos-Magdaleno M, Jutten RJ, et al. Assessing everyday activities across the dementia spectrum with the Amsterdam IADL Questionnaire. Curr Alzheimer Res. 2018;15:1261-6.
[29] Facal D, Juncos-Rabadan O, Guardia-Olmos J, Pereiro AX, Lojo-Seoane C. Characterizing Magnitude and Selectivity of Attrition in a Study of Mild Cognitive Impairment. J Nutr Health Aging. 2016;20:722-8.
[30] Molinuevo JL, Gramunt N, Gispert JD, Fauria K, Esteller M, Minguillon C, et al. The ALFA project: A research platform to identify early pathophysiological features of Alzheimer's disease. Alzheimers Dement (N Y). 2016;2:82-92.
[31] Dubois B, Epelbaum S, Nyasse F, Bakardjian H, Gagliardi G, Uspenskaya O, et al. Cognitive and neuroimaging features and brain beta-amyloidosis in individuals at risk of Alzheimer’s disease (INSIGHT-preAD): a longitudinal observational study. Lancet Neurol. 2018;17:335-46.
[32] Lee A, Alber J, Monast D, Menard W, Tang T, Bodge C, et al. The Butler Alzheimer’s Prevention Registry: recruitment and interim outcome. Alzheimers Dement. 2017;13:622-3.
[33] Stringer G, Brown L, Leroi I, Sikkes SAM. Capturing declining daily activity performance in a technologically-advancing older population: UK cultural validation of the Amsterdam IADL Questionnaire. British Society of Gerontology 45th Annual Conference. Stirling2016.
[34] Petrovic J, Milosevic V, Zivkovic M, Stojanov M, Milojkovic O, Kalauzi A, et al. Slower EEG alpha generation, synchronization and “flow”-possible biomarkers of cognitive impairment and neuropathology of minor stroke. PeerJ. 2017;5:e3839.
[35] Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009;5:27-48.
[36] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research. 1975;12:189-98.
[37] Huppert FA, Jorm AF, Brayne C, Girling DM, Barkley C, Beardsall L, et al. Psychometric properties of the CAMCOG and its efficacy in the diagnosis of dementia. Aging Neuropsychol C. 1996;3:201-14.
[38] Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int Psychogeriatr. 1997;9 Suppl 1:173-6.
[39] Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS) - Recent evidence and development of a shorter version. Clinical Gerontologist. 1986;5:165-73.
[40] Muthén LK, Muthén BO. Mplus User's Guide. 7th ed. Los Angeles, CA: Muthén & Muthén; 2012.
[41] Brown TA. Confirmatory factor analysis for applied research. 2nd ed. New York: The Guilford Press; 2015.
[42] Choi SW, Gibbons LE, Crane PK. lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. Journal of Statistical Software. 2011;39.
[43] Herrera AN, Gomez J. Influence of equal or unequal comparison group sample sizes on the detection of differential item functioning using the Mantel-Haenszel and logistic regression techniques. Qual Quant. 2008;42:739-55.
[44] De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. Cambridge: Cambridge University Press; 2011.
[45] Gelin MN, Zumbo BD. Differential item functioning results may change depending on how an item is scored: An illustration with the Center for Epidemiologic Studies Depression Scale. Educ Psychol Meas. 2003;63:65-74.
[46] Jodoin MG, Gierl MJ. Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. Applied Measurement in Education. 2001;14:329-49.
[47] IBM Corp. IBM SPSS Statistics for Windows. 22 ed. Armonk, NY: IBM Corp.; 2013.
[48] R Core Team. R: A language and environment for statistical computing. 3.5.3 ed2019.
[49] Fleishman JA, Spector WD, Altman BM. Impact of differential item functioning on age and gender differences in functional disability. J Gerontol B Psychol Sci Soc Sci. 2002;57:S275-84.
[50] Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, et al. 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. Med Care. 2004;42:I37-48.