

The influence of diversity on the measurement of functional impairment: An international validation of the Amsterdam IADL Questionnaire in eight countries

Mark A. Dubbelman¹ | Merike Verrijp¹ | David Facal² | Gonzalo Sánchez-Benavides³ | Laura J.E. Brown⁴ | Wiesje M. van der Flier^{1,5} | Hanna Jokinen^{6,7} | Athene Lee⁸ | Iracema Leroi⁴ | Cristina Lojo-Seoane² | Vuk Milošević⁹ | José Luís Molinuevo³ | Arturo X. Pereiro Rozas² | Craig Ritchie¹⁰ | Stephen Salloway⁸ | Gemma Stringer⁴ | Stelios Zygouris^{11,12} | Bruno Dubois¹³ | Stéphane Epelbaum¹³ | Philip Scheltens¹ | Sietske A.M. Sikkes^{1,5}

¹Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

²Department of Developmental Psychology, University of Santiago de Compostela, A Coruña, Spain

³BarcelonaBeta Brain Research Center (BBRC), Pasqual Maragall Foundation, Barcelona, Spain

⁴Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Science Centre, Manchester, UK

⁵Department of Epidemiology and Biostatistics, Amsterdam UMC, Amsterdam, The Netherlands

⁶Clinical Neurosciences, Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

⁷Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Finland

⁸Butler Hospital, Warren Alpert Medical School of Brown University, Providence, Rhode Island

⁹Clinic of Neurology, Clinical Center Niš, Niš, Serbia

¹⁰University of Edinburgh, Edinburgh, UK

¹¹School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

¹²Network Aging Research, Heidelberg University, Heidelberg, Germany

¹³Department of Neurology, Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A) of the Pitié-Salpêtrière Hospital & ARAMIS, Sorbonne University, Inria de Paris, Institut du cerveau et de la moelle épinière (ICM), Paris, France

Correspondence

Mark A. Dubbelman, Amsterdam UMC, Alzheimer Center, Department of Neurology, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands.
Email: m.dubbelman@amsterdamumc.nl

Funding information

Memorabel, Grant/Award Number: 733050205

Abstract

Introduction: To understand the potential influence of diversity on the measurement of functional impairment in dementia, we aimed to investigate possible bias caused by age, gender, education, and cultural differences.

Methods: A total of 3571 individuals (67.1 ± 9.5 years old, 44.7% female) from The Netherlands, Spain, France, United States, United Kingdom, Greece, Serbia, and Finland were included. Functional impairment was measured using the Amsterdam Instrumental Activities of Daily Living (IADL) Questionnaire. Item bias was assessed using differential item functioning (DIF) analysis.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* published by Wiley Periodicals, Inc. on behalf of Alzheimer's Association

Results: There were some differences in activity endorsement. A few items showed statistically significant DIF. However, there was no evidence of meaningful item bias: Effect sizes were low (ΔR^2 range 0-0.03). Impact on total scores was minimal.

Discussion: The results imply a limited bias for age, gender, education, and culture in the measurement of functional impairment. This study provides an important step in recognizing the potential influence of diversity on primary outcomes in dementia research.

KEYWORDS

Alzheimer's disease, cross-cultural validation, dementia, differential item functioning, diversity, functional decline, instrumental activities of daily living, item response theory

1 | INTRODUCTION

Impairment in cognitively complex “instrumental activities of daily living” (IADLs), such as doing grocery shopping, managing personal finances, and using mobile devices, may be one of the first symptoms of dementia.¹⁻³ 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.^{5,6} Therefore, functional impairment is an important and highly relevant outcome measure for clinical practice and clinical trials. In recently drafted industry guidelines, the U.S. Food and Drug Administration (FDA) recommended the use of functional impairment as a measure for effectiveness of treatment and of disease progression.⁷ It is a potential global outcome measure in dementia research.^{8,9}

Because everyday functioning relates to daily life, IADLs may be especially sensitive to bias caused by various factors, such as age, gender, and cultural differences. Previous studies have shown gender effects on traditional IADL instruments,⁹⁻¹² as they include predominantly household activities, which may be performed more often by women. Scientific literature concerning cultural and ethnographic diversity in the context of dementia is scarce.^{13,14} The selection of activities to include in an IADL instrument may be culture-specific. For example, in the United States, it is customary to write checks, whereas in The Netherlands, people often use online banking. Mere translation of an instrument does not always account for national (cross-cultural) disparities,^{15,16} and although many functional instruments have been translated into numerous languages, there is no gold standard for cross-cultural adaptation of questionnaires.¹⁷ This emphasizes the importance of investigating potential sources of bias and their influence on item and scale level.

We aimed to study the potential influences of diversity on the measurement of functional impairment using the Amsterdam IADL Questionnaire (A-IADL-Q). Specifically, we investigated item bias caused by various factors: cross-cultural differences (operationalized by using country of residence), age, gender, and education. We obtained data from eight Western countries: The Netherlands, Spain, France, United States, United Kingdom, Greece, Serbia, and Finland.

2 | METHODS

The present study included data from 3571 individuals with a completed A-IADL-Q from memory clinics and cognition studies from eight countries: The Netherlands (Amsterdam Dementia Cohort¹⁸ and European Prevention of Alzheimer's Dementia Longitudinal Cohort Study, EPAD^{19,20}), Spain (Compostela Aging Study^{21,22}; EPAD; and Alfa+ project²³), France (investigation of Alzheimer's predictors in subjective memory complainers (INSIGHT-preAD) study²⁴; EPAD; and Socrates study), United States (Butler Alzheimer's Prevention Registry²⁵), United Kingdom (EPAD and software architecture for mental health self management (SAMS) project²⁶), Greece (Greek Association for Alzheimer's Disease and Related Disorders), Serbia (Niš Clinic of Neurology²⁷), 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.

2.1 | Measures

2.1.1 | Amsterdam IADL Questionnaire (A-IADL-Q)

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

Unlike many other IADL instruments,²⁸ the A-IADL-Q has been validated extensively and has been shown to have good internal consistency, validity, and reliability.^{29–31} Furthermore, it appears to be independent of age and gender,³⁰ and sensitive to change over time.³² The short version was developed to create a more concise measure, as well as to reduce potential cultural bias by only including widely relevant activities.³³ International use of the A-IADL-Q is steadily increasing. All translations have gone through a cross-cultural adaptation process based on procedures described by Beaton et al.³⁴ in which experts and prospective users were asked to evaluate the translated instrument (a more detailed description of this process can be found in the Supplementary Material).

The questionnaire is scored using item response theory (IRT), as described elsewhere.^{29,31} IRT assumes that an instrument measures a latent trait, which is represented in a scale ranging from total absence to abundance of the particular trait.³⁵ The A-IADL-Q latent trait is “IADL functioning.”³⁰ In IRT, parameters are calculated for each item, which contain information about item response category location (or difficulty, ie, at which trait level half the population endorses a given response category of an item), as well as slope (or discriminatory ability, ie, how well an item can distinguish between people with lower and higher levels of the trait).

All A-IADL-Q 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 representing the trait level 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

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). Finally, 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 were not obtained for all included participants: We excluded individuals living in nursing homes ($n = 130$) because they have limited IADL independence.

2.2 | Statistical analyses

We investigated item bias using “differential item functioning” (DIF) analysis. DIF analysis is a technique for identifying items that have different item locations and/or slopes in different groups. DIF is assumed to occur when the relationship between a test item and the latent trait

HIGHLIGHTS

- Diversity in age, gender, education, and culture may influence measurement of instrumental activities of daily living (IADLs).
- A total of 3571 people from eight countries answered the Amsterdam IADL Questionnaire (A-IADL-Q).
- Minor item bias was found for country, with a marginal influence on total scores.
- No meaningful item bias was found for age, gender, and education.
- These findings provide evidence for valid measurement of everyday functioning.

RESEARCH IN CONTEXT

- **Systematic review:** Instrumental activities of daily living (IADLs) are important for the diagnosis of dementia and may form an important outcome measure for clinical trials. IADL measures are widely used internationally. We reviewed almost 200 records from a PubMed search of cross-cultural comparability of IADLs in dementia and found only a handful of publications targeting cross-cultural validation, mostly into a single language.
- **Interpretations:** Our findings of absence of meaningful bias for cultural differences, age, gender, and education in the Amsterdam IADL Questionnaire (A-IADL-Q) suggest that the A-IADL-Q is a suitable instrument for the measurement and international comparison of decline in IADL functioning in early dementia in a demographically diverse population.
- **Future directions:** Development of new translations of the A-IADL-Q can make the instrument truly globally applicable. In addition, ethn racial disparities should also be taken into account in future research so that IADLs can be used as a universal measure.

is not the same across study-irrelevant groups.³⁵ It is considered a variation in measurement and is therefore undesirable.⁴⁰ We studied DIF in the following groups: (1) nationality, using the Dutch cohort as a reference group, while grouping all other studies by country; (2) men and women; and, based on median split; (3) young (<67.2 years) and old age (≥ 67.2 years); and (4) low (<12 years) and high education (≥ 12 years).

For all DIF analyses, a minimum count of one case in at least two different response categories was required in each group for every item. We used the ordinal logistic regression (OLR) approach, which is often used and can be performed in standard software. OLR has been shown

TABLE 1 Information about participants, inclusion and exclusion criteria, and information about the A-IADL-Q administration per included sample

Study name	Country	Amsterdam Dementia Cohort ¹⁸	European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD) ^{19,20}	Alfa+ Study ²³	INSIGHT preAD ²⁴	Butler Alzheimer's Prevention Registry ²⁵	Greek Association of Alzheimer's Disease and Related Disorders	Niš Clinic of Neurology ²⁷	Helsinki Small Vessel Disease study	SAMS Project ²⁶
	Netherlands	Spain	Spain (n = 218) France (n = 103) Netherlands (n = 88) United Kingdom (n = 71)	Spain	France	United States of America	Greece	Serbia	Finland	United Kingdom
Participants included	1429	600	480	333	308	154	61	45	43	22
Age range	25-84 years	50-101 years	51-88 years	49-73 years	70-85 years	58-77 years	65-92 years	26-93 years	66-75 years	65-82 years
Research environment	Consecutive memory clinic patients	MCI patients referred by GP	Participants from existing study cohorts	Mostly offspring of AD patients	Consecutive memory clinic patients and advertisement recruited	Advertisement recruited	Patients from day center for dementia	Memory clinic patients	Patients with neuroimaging data selected from existing databank	Recruited from dementia research registry, memory clinic patients
Relevant inclusion and exclusion criteria	None	Cognitive complaints without dementia; Age ≥ 50 years	No dementia; Age ≥ 50 years	CN (MMSE ≥ 26, CDR0); No neurological diseases; Age 45-74 years	CN (MMSE ≥ 27, CDR0); Amyloid PET at baseline; No episodic memory deficits, no neurological diseases; Not living in nursing home; Age 70-85 years	CN or mild memory loss; No neurological diseases or dementia diagnosis; Age 55-85 years	Dementia-related diagnosis; Reliable informant; No neurological diseases other than dementia; Age ≥ 65 years	CN, MCI, post-stroke cognitive impairment	No major neurological symptoms or psychiatric disease; Independence in basic ADL; No large infarcts, hemorrhages, contusion or tumor on MRI; Age 65-75 years	SCD (ECog ≥ 1436 and answered "yes" when asked if "concerned they have a memory or other thinking problem"), MCI; Age ≥ 65 years
A-IADL-Q version	Original (n = 730) SV (n = 699)	Original	Original	SV	Original	SV	SV	SV	SV	Version adapted from original
Clinical measures										
Participants (%) selected for validation ^a	1369 (95.8)	300 (50.0)	480 (100.0)	333 (100.0)	308 (100.0)	154 (100.0)	26 (42.6)	45 (100.0)	43 (100.0)	22 (100.0)
Measures available	MMSE, CAMCOG, CDR, GDS	MMSE, CAMCOG, GDS	MMSE, CDR, GDS	MMSE, CDR	MMSE, CDR	MMSE, GDS	MMSE, CAMCOG, CDR, GDS	MMSE	MMSE, GDS	GDS

Additional inclusion and exclusion criteria are available upon request.

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

^aParticipants living in a nursing home were excluded from validation and no clinical measures were obtained for them.

TABLE 2 Demographics and clinical characteristics for all participants, and grouped per country

	All	The Netherlands	Spain	France	United States	United Kingdom	Greece	Serbia	Finland
Total n	3571	1515	1151	509	154	93	61	45	43
Females, n (%) ^a	1597 (44.7)	637 (42.0)	485 (42.1)	262 (51.5)	104 (67.5)	43 (46.2)	18 (29.5)	25 (55.6)	23 (53.5)
Age (years)	67.14 ± 9.5	63.78 ± 8.5	67.84 ± 10.4	73.48 ± 6.2	66.65 ± 4.5	68.42 ± 5.8	79.99 ± 6.4	65.44 ± 13.1	71.69 ± 2.8
Education years	12.19 ± 3.9	11.34 ± 3.2	11.97 ± 4.4	13.95 ± 3.7	16.82 ± 2.3	12.99 ± 3.1	9.50 ± 4.3	13.93 ± 4.3	12.93 ± 5.5
Dementia diagnosis, n (%) ^a	860 (29.9)	647 (47.2)	188 (20.2)	0 (0)	0 (0)	0 (0)	21 (80.8)	4 (8.9)	0 (0)
A-IADL-Q									
T-score ^b	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
Clinical measures^a									
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
CAMCOG	78.57 ± 17.3	78.75 ± 16.1	80.98 ± 19.1	—	—	—	41.62 ± 9.7	—	—
CDR, M (IQR)	0 (0–0.5)	0.5 (0–1)	0 (0–0)	0 (0–0)	—	0 (0–0)	2 (0.5–2)	—	—
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

All data are displayed as mean ± standard deviation, except as stated otherwise. “—” denotes that data were not available.

A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; CAMCOG, Cambridge Cognitive Examinations; CDR, Clinical Dementia Rating; GDS, Geriatric Depression Scale; IQR, interquartile range; M, median; MMSE, Mini-Mental State Examination.

^aData were not obtained for all participants.

^bThe score shown is based on either the original or short version of the A-IADL-Q, as administered to each participant.

previously to be superior the Mantel-Haenszel procedure.⁴¹ We used the “lordif” package version 0.3-3 for R, developed by Choi et al⁴⁰; “lordif” has been used extensively in the literature, ensuring appropriateness and replicability of our procedures. In the OLR approach, a null model and three hierarchically nested models are created and compared for each item. When DIF is present and constant across all levels of the latent trait, it is called uniform DIF. The response categories of an item with uniform DIF are located at a different location in each group.⁴² When an item is easier at one level of the trait and more difficult at another level, it is considered to have non-uniform DIF.⁴² Items with non-uniform DIF have different discriminatory abilities in each group. Statistically significant DIF was determined on the basis of the likelihood-ratio χ^2 test with an α level of .01, to avoid type I error, and because multiple nested models are being tested for each item. Because of inflated type I error in OLR DIF analyses,⁴³ we added a step to establish presence of practically meaningful DIF,^{44,45} based on a McFadden's pseudo R^2 (ΔR^2) value of .035 or larger. This approach reduces the risk of finding significant but negligible DIF, albeit at the cost of a reduction in power.⁴³ Furthermore, we used the following effect size criteria to quantify DIF size: ΔR^2 values between .035 and .070 for moderate, and above .070 for large DIF.⁴³ To refine DIF detection and effect size estimates, we then performed Monte Carlo simulations over 1000 replications in which the detection criteria as well as effect size measures are computed repeatedly over simulated data based on the empirical data sets. The simulated data are generated under the hypothesis that there is no DIF, while keeping the observed group differences in trait levels.

As a means of construct validation, Pearson's r for continuous or Kendall's τ 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).

Data were processed in SPSS Statistics version 22⁴⁶ and R version 3.6.1.⁴⁷

3 | RESULTS

On average, participants were 67.1 ± 9.5 (m ± SD) years old. Table 2 shows the demographics and clinical measures of all participants, stratified by country.

The overall mean score on the A-IADL-Q was 58.40 ± 14.2 . A-IADL-Q scores per country are shown in Table 2.

3.1 | Item endorsement

Generally, item endorsement was comparable between countries, as well as between men and women, younger and older participants, and participants with lower and higher education. Table 3 highlights a few activities in which there were apparent differences. “Minor repairs” was endorsed by a larger percentage of men, as compared to women. Conversely, “using a washing machine” was endorsed more often by women. Participants with a lower indication endorsed “withdrawing cash from an ATM” somewhat less often than participants with a higher education. Older participants were less likely to work, compared to younger participants. Participants from Greece, Spain, and Serbia used computers less often than participants from the other countries. Participants from the United States appeared to use public transportation less often than participants from European countries (see Table 3).

TABLE 3 Differences in endorsement in selected activities

Activity	Country					Age			Gender		Education			
	The Netherlands	Spain	France	United States	United Kingdom	Greece	Serbia	Finland	Young	Old	Men	Women	Low	High
Minor repairs	46.2%	57.9%	67.4%	55.8%	62.4%	55.7%	57.8%	72.1%	53.6%	55.8%	70.9%	42.4%	52.8%	58.8%
Washing machine	58.4%	70.7%	77.0%	92.2%	75.3%	63.9%	71.1%	81.4%	72.4%	63.5%	44.0%	93.1%	65.7%	73.1%
Withdrawing cash from ATM	69.6%	64.7%	82.7%	74.0%	80.6%	9.8%	55.6%	72.1%	77.2%	62.6%	75.8%	75.4%	66.2%	82.0%
Working	52.3%	42.4%	54.2%	66.9%	24.7%	9.8%	53.3%	58.1%	61.6%	36.2%	53.1%	50.9%	47.4%	55.4%
Using a computer	82.0%	52.7%	75.2%	97.4%	94.6%	8.2%	53.3%	81.4%	82.1%	60.3%	79.7%	73.3%	65.3%	84.3%
Public transportation	49.7%	70.5%	86.2%	27.9%	59.1%	83.6%	51.1%	79.1%	58.0%	64.5%	59.0%	66.5%	55.4%	68.0%

Differences of interest between groups within each factor (country, age, gender, and education) are displayed in bold. Endorsement of other activities included in the Amsterdam IADL Questionnaire did not differ as much and these activities are not displayed here.

3.2 | Item bias

Due to restricted variability in some items, we were unable to analyze all items. Two hundred seventy-two of 300 items (90.7%) in the A-IADL-Q-SV were analyzed. Of the items analyzed, 26.6% had statistically significant DIF. Effect sizes were very small for all factors (ΔR^2 range .000-.034, see Figure 1). Monte Carlo simulations showed that the mean p -value for the χ^2 statistic across all items varied between comparisons from .006 to .012, which was close to the .01 α -level used to detect DIF. Simulation-based thresholds for effect size ranged from .001 to .018 across all analyses (Figure 1). Lowering of the threshold would lead to more items being flagged for DIF. The effect sizes, however, remained very small.

For the original version, 437 of 490 items (89.2%) were analyzed. Of those, 20.4% had statistically significant DIF. The effects for age, gender, and education were again small (ΔR^2 range .000-.032). Four items showed meaningful DIF for nationality with a moderate effect. In Spain, “using the washing machine” ($\Delta R^2 = .043$), “making appointments” ($\Delta R^2 = .064$), and “playing card and board games” ($\Delta R^2 = .043$) were flagged. All three items had uniform DIF: 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” ($\Delta R^2 = .064$). 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, thereby correcting for the effect of DIF. In the Spanish group, the mean score decreased by .16 points on the T -scale, in the French group, the mean score decreased by .07 points. The largest individual differences in both countries (-1.14 and -1.33 , respectively) corresponded to a difference of approximately one tenth of an SD, and can therefore be considered negligible. Figure 2 shows the individual score changes after DIF correction in Spain and France. There was no meaningful bias for nationality in the other countries. Simulations showed the mean χ^2 statistic p -value across all items varied from .008 to .012. The largest ΔR^2 effect size was .026 (range .001-.026), which corresponds to a negligible effect.

3.3 | A-IADL-Q-SV construct validation

Overall, all correlations were in the same directions and of similar magnitudes as compared to the original validation data from The Netherlands.³⁰ Age seemed more strongly associated with IADL impairment in Spain ($r = -0.47$, 95% confidence interval [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

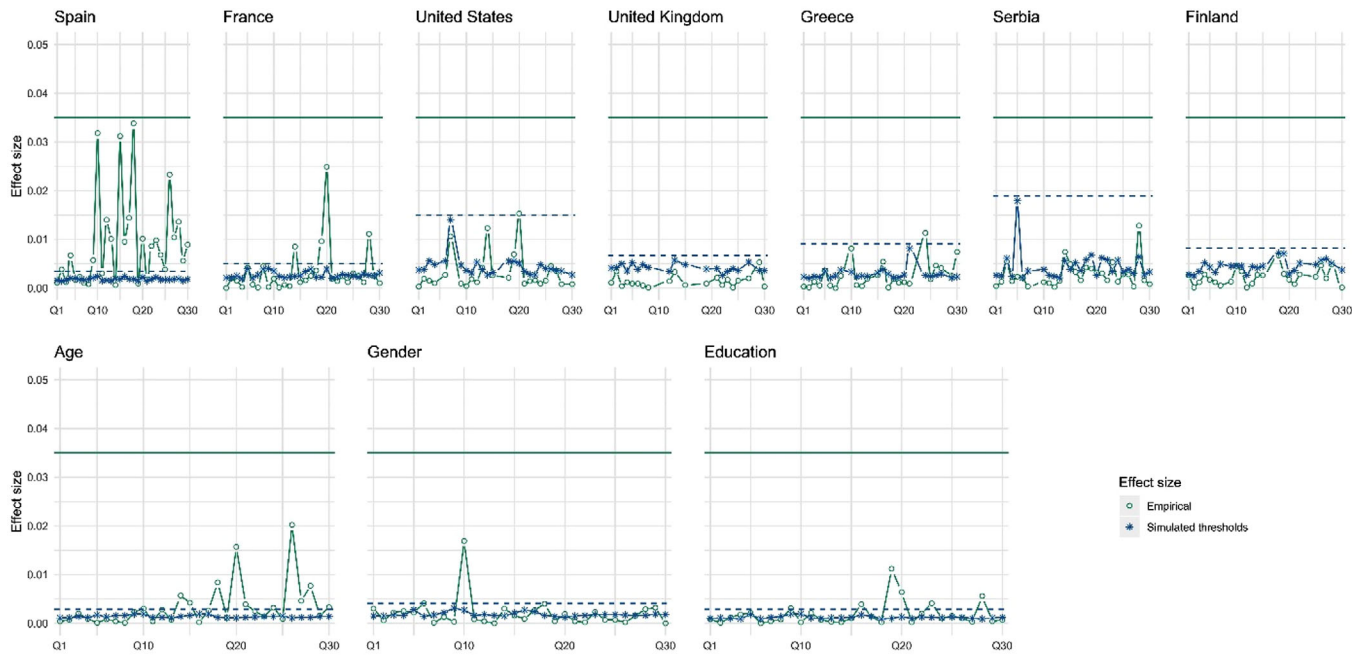


FIGURE 1 DIF effect sizes for country, age, gender, and education in the A-IADL-Q-SV. Green circles represent the empirically found ΔR^2 effect sizes; blue asterisks represent the 99th percentile ΔR^2 effect sizes from MC simulations. A solid green line is placed at the predetermined threshold for practically meaningful DIF ($\Delta R^2 = .035$); a dashed blue line is placed just above the highest simulated effect size threshold. Abbreviations: A-IADL-Q, Amsterdam IADL Questionnaire; DIF, differential item functioning

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 Supplementary Material.

4 | DISCUSSION

In this study, we demonstrated that the influence of diversity on the measurement of IADL impairment, as measured with the A-IADL-Q, seems minimal. Although we found some differences with regard to activity endorsement between countries, there was no evidence of practically meaningful item bias caused by various factors, including age, gender, education, and culture. These findings, together with the similar associations with demographic, cognitive, and functional measures as found in earlier validation efforts,³⁰ further support the validity of the A-IADL-Q.

Addressing potential bias caused by various types of diversity is highly relevant in dementia research.¹⁴ With respect to the measurement of functional impairment, there have been contradictory findings, with some studies showing a general comparability of IADLs across cultures and different ethnoracial groups,^{8,9} and others reporting differences between cultures, genders, and ages.⁴⁸⁻⁵¹ For an optimal comparison of functional outcome in international studies and clinical trials, a valid, cross-culturally adapted instrument is crucial. In the present study, the relevance of addressing potential bias was underscored by the fact that we found some differences in activity

endorsement, particularly in activities related to the household and to technology. Gender roles can differ between countries, and they might determine the IADL activities in which one participates. In Mediterranean countries, it seemed people used computers less often than in Northern European countries and America.

In our current sample, the effects of DIF were small and thus did not pass our threshold for practically meaningful DIF. The reason that we found little evidence of meaningful DIF may be attributed to the cross-cultural adaptation process that all translations went through, in which potential cross-cultural differences were identified beforehand and cultural adaptations were made as necessary. These changes were minor, and we believe the items included should be applicable to Western culture in general. As part of the development of the short version, international experts provided feedback on the cross-cultural comparability of the items,³³ which may explain the absence of practically meaningful item bias for nationality. Because the A-IADL-Q-SV does not appear to have practically meaningful item bias, *T*-scores do not need to be adjusted to be compared across countries, ages, genders, or levels of education. This suggests that the A-IADL-Q yields valid and cross-culturally comparable estimations of functional decline. Previous studies^{30,31,33} have already shown that A-IADL-Q scores are independent of age, gender, and education, and our findings corroborate this. This is an important finding, because other functional instruments do appear to be biased for gender, age, and cultural differences.^{48,49}

In the original version, a few items appeared to be biased in Spain and France. “Making appointments” had the largest DIF effect, and

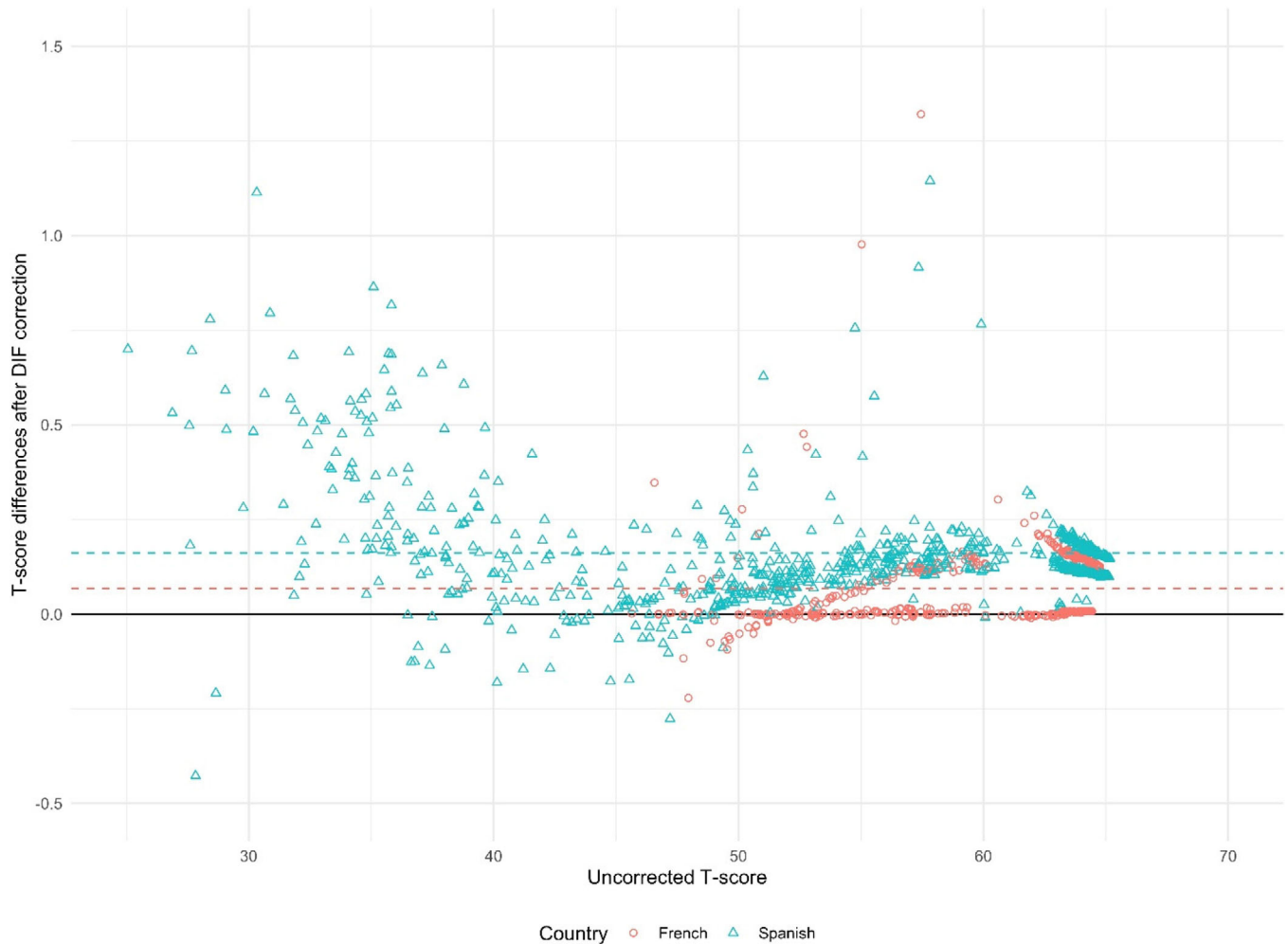


FIGURE 2 Scatter plot showing the differences between initial (uncorrected) and DIF-corrected T-scores for the A-IADL-Q in the French (red) and Spanish (blue) groups, plotted against the uncorrected T-scores. A dashed line is placed at the mean change in score in the French and Spanish groups. Difference in total score ranges from -0.5 to $+1.5$ on the T-score, corresponding to approximately one tenth of a standard deviation difference. A solid black line is placed at no change. Abbreviations: A-IADL-Q, Amsterdam IADL Questionnaire; DIF, differential item functioning

a potential explanation 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,” whereas the intended definition was broader. However, adding examples may actually have restricted the interpretation of the question to the specific examples given, and led to a loss of the broader meaning. 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.

The associations between A-IADL-Q-SV scores and demographic, cognitive, and functional measures we found here largely correspond to those previously described for the original version.³⁰ 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.²¹ In France, the United States, and the United Kingdom, the studies recruited mainly cognitively healthy

participants, resulting in 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 into the impact of the bias on the overall scores and it allows for correction. We additionally used simulations to further validate the empirical findings. 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 3500 individuals from 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 included data from only eight developed, Western countries. Our findings cannot be generalized to other parts of the world. One study found DIF in an IADL instrument between different Asian cultures.⁵² It should also be noted that we use the term “culture” to refer to each country’s national culture. Furthermore, we did not have access to information about ethnicity or race. It is currently unclear what the influence of ethnora- cial differences are on the measurement of IADLs. Second, our sample comprised mainly highly educated people. The group we defined as having low education still received up to 12 years of education. It is possible that different results would be obtained in samples with less formal education. Third, the sample size was relatively small in Finland, Serbia, and Greece. This may have reduced our power to detect DIF. We tried to address this issue by performing Monte Carlo simulations, which indicated that the predetermined cutoff for practically meaningful DIF may have been somewhat high. More items would show DIF, if the threshold was lowered. However, when considering how these findings influence the total score, the impact seems minimal and the DIF effect sizes remain small.

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. More research is needed to understand the differences between Western and Oriental and other cultures, as well as differences between ethnicities and races.

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 seem to have meaningful item bias, has good construct validity, and is more pragmatic.

To conclude, we found no indication of the presence of clinically relevant bias caused by several aspects of diversity, including age, gender, education, and cultural differences. This is important, because it further underscores 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.

ACKNOWLEDGMENTS

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 present study is supported by a grant from Memorabel (733050205), which is the research program of the Dutch Deltaplan for Dementia. The chair of WMF is supported by the Pasman stichting. The clinical database structure for the Amsterdam Dementia Cohort was developed with funding from Stichting Dioraphte. DF, CLB and AXPR are supported by FEDER grant PSI2014-55316-C3-1-R, the Spanish National Research Agency grant PSI2017-89389-C2-1-R and the Galician Government GI-1807-USC: Ref. ED431-2017/27. GSB, JLM and the ALFA+ project

has received funding from “la Caixa” Foundation (ID 100010434), under agreement LCF/PR/GN17/50300004 and the Alzheimer’s Association and an international anonymous charity foundation through the TriBEKa Imaging Platform project (TriBEKa-17-519007). CWR and the work for EPAD has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking EPAD grant agreement n° 115736. SE is supported by a joint collaborative grant by the AP-HP and Inria. LJEB, IL and GS were supported by the U.K. Engineering and Physical Sciences Research Council #EP/K015796/1. SZ is supported by the Robert Bosch Foundation Stuttgart within the Graduate Program People with Dementia in General Hospitals, located at the Network Aging Research (NAR), Heidelberg University, Germany. VM is supported by the Serbian Ministry of Education, Science and Technological Development, grant OI 173022. AL is partially supported by Institutional Development Award Number U54GM115677 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds Advance Clinical and Translational Research (Advance-CTR). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The other authors did not receive funding directly related to this work.

CONFLICT OF INTEREST

The Amsterdam IADL Questionnaire was developed by S.A.M.S. and P.S., who were involved in the conception of the present study. The other authors report no conflict of interests.

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(3):179-186.
3. Pérès K, Helmer C, Amieva H, 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(1):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(1):63-71.
5. Luck T, Riedel-Heller SG, Luppá M, et al. A hierarchy of predictors for dementia-free survival in old-age: results of the AgeCoDe study. *Acta Psychiatr Scand*. 2014;129(1):63-72.
6. Weintraub S, Carrillo MC, Farias ST, et al. Measuring cognition and function in the preclinical stage of Alzheimer’s disease. *Alzheimers Dement*. 2018;4:64-75.
7. 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.
8. Truscott DJ. Cross-cultural ranking of IADL skills. *Ethn Health*. 2000;5(1):67-78.
9. Nikula S, Jylhä M, Bardage C, et al. Are IADLs comparable across countries? Sociodemographic associates of harmonized IADL measures. *Aging Clin Exp Res*. 2003;15(6):451-459.
10. 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.

11. Lechowski L, de Stampa M, Denis B, 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(1):46-53.
12. Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? *Neurodegener Dis Manag*. 2014;4(5):393-400.
13. 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(3):187-195.
14. Babulal GM, Quiroz YT, Albeni BC, Arenaza-Urquijo EM. Perspectives on ethnic and racial disparities in Alzheimer's disease and related dementias: Update and areas of immediate need. *Alzheimers Dement*. 2018. <https://doi.org/10.1016/j.jalz.2018.09.009>.
15. Parra MA. Overcoming barriers in cognitive assessment of Alzheimer's disease. *Dement Neuropsychol*. 2014;8(2):95-98.
16. 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(1):116-128.
17. 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(4):435-441.
18. van der Flier WM, Scheltens P. Amsterdam dementia cohort: performing research to optimize care. *J Alzheimers Dis*. 2018;62(3):1091-1111.
19. Ritchie CW, Molinuevo JL, Truyen L, Satlin A, Van der Geysen S, Lovestone S, European Prevention of Alzheimer's Dementia (EPAD) Consortium. Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. *Lancet Psychiatry*. 2016;3(2):179-186.
20. Solomon A, Kivipelto M, Molinuevo JL, Tom B, Ritchie CW, the EPAD Consortium. European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol. *BMJ Open*. 2018;8(12). <https://doi.org/10.1136/bmjopen-2017-021017>.
21. Facal D, Carabias MAR, Pereiro AX, et al. Assessing everyday activities across the dementia spectrum with the Amsterdam IADL Questionnaire. *Curr Alzheimer Res*. 2018;15(13):1261-1266.
22. Facal D, Juncos-Rabadán 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(7):722-728.
23. Molinuevo JL, Gramunt N, Gispert JD, et al. The ALFA project: a research platform to identify early pathophysiological features of Alzheimer's disease. *Alzheimers Dement (N Y)*. 2016;2(2):82-92.
24. Dubois B, Epelbaum S, Nyasse F, 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(4):335-346.
25. Lee A, Alber J, Monast D, et al. The Butler Alzheimer's Prevention Registry: recruitment and interim outcome. *Alzheimers Dement*. 2017;13(7):622-623.
26. Stringer G, et al. Capturing declining daily activity performance in a technologically-advancing older population: UK cultural validation of the Amsterdam IADL Questionnaire. In British Society of Gerontology 45th Annual Conference. 2016. Stirling.
27. Petrovic J, et al. Slower EEG alpha generation, synchronization and "flow"-possible biomarkers of cognitive impairment and neuropathology of minor stroke. *PeerJ*. 2017;5:e3839. <https://doi.org/10.7717/peerj.3839>.
28. Sikkes SAM, de Lange-de Klerk ESM, Pijnenburg YAL, Scheltens P, Uitdehaag BMJ. A systematic review of Instrumental Activities of Daily Living scales in dementia: room for improvement. *J Neurol Neurosurg Psychiatry*. 2009;80(1):7-12.
29. Sikkes SAM, de Lange-de Klerk ESM, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. *Alzheimers Dement*. 2012;8(6):536-543.
30. Sikkes SAM, Knol DL, Pijnenburg YAL, de Lange-de Klerk ESM, Uitdehaag BMJ, Scheltens P. Validation of the Amsterdam IADL Questionnaire(c), a new tool to measure instrumental activities of daily living in dementia. *Neuroepidemiology*. 2013;41(1):35-41.
31. Sikkes SAM, Pijnenburg YAL, Knol DL, de Lange-de Klerk ESM, Scheltens P, Uitdehaag BMJ. 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(4):244-250.
32. Koster N, Knol DL, Uitdehaag BMJ, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). *Alzheimers Dement*. 2015;11(10):1231-1240.
33. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: development and validation of a short version of the Amsterdam IADL Questionnaire. *Alzheimers Dement*. 2017;8:26-35.
34. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;25(24):3186-3191.
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. *J Psychiatr Res*. 1975.12(3):189-198.
37. Huppert FA, Jorm AF, Brayne C, et al. Psychometric properties of the CAMCOG and its efficacy in the diagnosis of dementia. *Aging Neuropsychol Cogn*. 1996;3(3):201-214.
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):173-176.
39. Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS) - Recent evidence and development of a shorter version. *Clin Gerontol*. 1986;5(1-2):165-173.
40. 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. *J Stat Softw*. 2011;39(8):1-30.
41. 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. *Quality & Quantity*. 2008;42(6):739-755.
42. De Vet HCW, et al. *Measurement in Medicine*. Cambridge: Cambridge University Press; 2011.
43. Jodoin MG, Gierl MJ. Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Appl Measure Educ*. 2001;14(4):329-349.
44. 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 Measure*. 2003;63(1):65-74.
45. Rouquette A, Hardouin J-B, Vanhaesebrouck A, Sébille V, Coste J. Differential item functioning (DIF) in composite health measurement scale: Recommendations for characterizing DIF with meaningful consequences within the Rasch model framework. *PLoS One*. 2019;14(4):e0215073.
46. IBM Corp. *IBM SPSS Statistics for Windows*. Armonk, NY: IBM Corp.; 2016.
47. R Core Team. *R: A Language and Environment for Statistical Computing*, Vienna, Austria. 2019. <https://www.R-project.org/>.
48. 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(5):S275-S284.
49. Tennant A, Penta M, Tesio L, 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(1 suppl):I37-I48.

50. Berezuk C, Zakzanis KK, Ramirez J, et al. Functional reserve: experience participating in instrumental activities of daily living is associated with gender and functional independence in mild cognitive impairment. *J Alzheimers Dis.* 2017;58(2):425-434.
51. Sheehan C, Domingue BW, Crimmins E. Cohort trends in the gender distribution of household tasks in the United States and the implications for understanding disability. *J Aging Health.* 2019;31(10):1748-1769.
52. Niti M, Ng T-P, Chiam P-C, Kua E-H. Item response bias was present in instrumental activity of daily living scale in Asian older adults. *J Clin Epidemiol.* 2007;60(4):366-374.

How to cite this article: Dubbelman MA, Verrijp M, Facal D, et al. The influence of diversity on the measurement of functional impairment: An International validation of the Amsterdam IADL Questionnaire in eight countries. *Alzheimer's Dement.* 2020;12:e12021.
<https://doi.org/10.1002/dad2.12021>