Health-Related Quality-of-Life Assessment in Dementia: Evidence of Cross-Cultural Validity in Latin America

Kia-Chong Chua
King's College London, and South London and Maudsley NHS
Foundation Trust, London, United Kingdom

Jan R. Böhnke
University of Dundee

Martin Prince
King's College London

Sube Banerjee
University of Sussex

All health-related quality-of-life (HRQL) measures for dementia have been developed in high-income countries and none were validated for cross-cultural use. Yet, the global majority of people living with dementia reside in low- and middle-income countries. We therefore investigated the measurement invariance of a set of self- and informant-report HRQL measures developed in the United Kingdom when used in Latin America. Self-reported HRQL was obtained using (DEMQOL) at a memory assessment service in the United Kingdom ($n = 868$) and a population cohort study in Latin America ($n = 417$). Informant reports were collected using DEMQOL-Proxy at both sites ($n = 909$ and $n = 495$). Multiple-group confirmatory bifactor models for ordered categorical item responses were estimated to evaluate measurement invariance. Results support configural, metric, and scalar invariance for the concept of general HRQL in DEMQOL and DEMQOL-Proxy. The dominant impact of general HRQL on item responses was evident across U.K. English and Ibero American Spanish versions of DEMQOL ($\omega_h = 0.87$–0.90) and DEMQOL-Proxy ($\omega_h = 0.88$–0.89). Ratings of "positive emotion" did not show a major impact on general HRQL appraisal, particularly for Latin American respondents. Item information curves show that self- and informant-reports were highly informative about the presence rather than the absence of HRQL impairment. We found no major difference in conceptual meaning, sensitivity, and relevance of DEMQOL and DEMQOL-Proxy for older adults in the United Kingdom and Latin America. Further replication is needed for consensus over which HRQL measures are appropriate for making cross-national comparisons in global dementia research.

Public Significance Statement
Measuring health-related quality-of-life (HRQL) in older adults living with dementia is important to quantify their burden of disease. Globally more than half of these adults live in low- and middle-income countries, but all HRQL measures for dementia originate from high-income countries. Our study is the first to show for a self- and informant-report HRQL instrument (DEMQOL) that the appraisals are comparable across U.K. (English) and Latin American (Spanish) community samples.

Keywords: dementia, health-related quality-of-life, cross-cultural validation, measurement invariance

Supplemental materials: http://dx.doi.org/10.1037/pas0000743.supp

According to the 2010 World Alzheimer Report (Wimo & Prince, 2010), the likelihood of developing dementia roughly doubles every 5 years after the age of 65. As the number of people reaching the age of 65 is growing rapidly due to population ageing worldwide, the number of people living with dementia is expected to rise globally (Prince, Guerchet, & Prina, 2015).

Dementia brings about a decline in memory, reasoning, and communication skills, and a gradual loss of skills needed in daily life for independent living (Knapp et al., 2007). At any stage of illness, individuals may also develop behavioral and psychological symptoms of dementia such as depression, psychosis (hallucinations and delusions), aggression, and wandering. Available drug treatment may improve symptoms temporarily, but none has been shown to slow or stop the disease process (The Alzheimer's Association, 2013). Current standard treatments continue to be the subject of clinical trials due to long-standing concerns over drug efficacy and safety (Ballard et al., 2009; Banerjee et al., 2011). With prevailing challenges in the treatment of dementia, the goal of "adding years to life" also needs to consider the goal of "adding life to years" (Clark, 1995).

Cognitive functioning is fundamentally a core outcome of disease-modifying treatment in dementia (Webster et al., 2017). However, interventions whose efficacy is tested on change in standardized cognitive test performance may not capture outcomes of greatest relevance to the lived experience of people with dementia (Harrison, Noel-Storr, Demeyere, Reynish, & Quinn, 2016). The goal of "adding life to years" needs an examination of dementia's impact on the whole person. This is the purpose of health-related quality-of-life (HRQL) measures (Dichter, Schwab, Meyer, Bartholomeyczik, & Halek, 2016).

Despite the clarity of purpose, the absence of a theoretical framework unique for HRQL in dementia has resulted in the emergence of at least 18 measures over the past 20 years (Missotten, Dupuis, & Adam, 2016). This diversity prompted calls for consensus over what should be the standard measurement tool for HRQL in dementia. Under the Core Outcome Measures in Effectiveness Trials initiative, efforts to establish evidence-based consensus on measurement tools have focused on community care settings for people with dementia (Reilly et al., 2014), as well as for disease modification trials in mild-to-moderate dementia (Webster et al., 2017).

A key motive for encouraging use of a standard HRQL measure across evaluation purposes is the need for making direct comparison between different types of dementia care interventions that incur disparate amount of resources to improve psychosocial outcomes (e.g., Cooper et al., 2012; Knapp, Iemmi, & Romeo, 2013; Spijker et al., 2008). Common use of a standard HRQL measure can also enhance interprofessional communication in clinical care (Bentvelzen, Aerts, Seeher, Wesson, & Brodaty, 2017). However, even with standard outcome measures in dementia like the Mini-Mental State Examination (MMSE), factors like ethnicity can distort measurements when there are no genuine differences (Dai et al., 2013; Jones, 2006). The use of a standard HRQL measure with diverse population needs investigation of such measurement issues.

In disease modification trials, where HRQL is increasingly considered for secondary outcome monitoring (Harrison et al., 2016), the need for meta-analyses to determine the overall benefit of treatment regimens is likely to involve comparisons of multiple clinical trials from different countries (e.g., Perng, Chang, & Tzang, 2018). Here, potential measurement issues due to the use of a standard HRQL measure with diverse populations may be accentuated in such international comparisons.

With increasing application across broader settings in diverse cultures, HRQL measurement in dementia faces a uniquely urgent challenge. From 2015 to 2050, the number of people living with dementia is predicted to increase about twofold in Europe and North America, threefold in Asia, and fourfold in Latin America and Africa (Prince, Guerchet, et al., 2015). More than half the world population of people with dementia currently live in regions classified by the World Bank as low- and middle-income countries and by 2050 this proportion is expected to rise to close to 70% due to population ageing (Prince, Wimo, et al., 2015). Although the global majority of people living with dementia resides in low- and middle-income countries, all HRQL measures for dementia have been developed in high-income countries and none is sufficiently validated to support use in cultures other than that where the original development took place (Dichter et al., 2016).

To date, among dementia-specific HRQL measures, the DEMQOL system (Mulhern et al., 2013) has the best evidence of responsiveness to minimum clinically important difference in cognitive function, behavioral and psychological symptoms in dementia, functioning in activities of daily living, and depression (Bentvelzen et al., 2017). Recent reviews have also consistently identified the DEMQOL system as among the most commonly used HRQL measures for dementia intervention and disease-modifying trials (Harrison et al., 2016; Webster et al., 2017). With self- and informant-report versions (DEMQOL and DEMQOL-Proxy), the DEMQOL system also highlight the need for comparing both perspectives to determine the utility of proxy report, especially for later stages of illness when self-report is often not available.

The prospect of standardizing HRQL measurement in dementia will need attention on measurement validity across cultures (Prince, 2008). We therefore used a unique dataset of HRQL assessments using DEMQOL and DEMQOL-Proxy, dementia-specific HRQL measures developed in the United Kingdom (Mulhern et al., 2013), to evaluate and cross-culturally validate its use in the United Kingdom, the Dominican Republic, Mexico, Cuba, Peru, and Venezuela.

Method

Study Participants

We conducted secondary data analysis on completely deidentified data drawn from two primary studies that had obtained ethics approval (Banerjee et al., 2007; Prince et al., 2007). The first comprised community-dwelling older adults attending a London memory assessment service. This is a diagnostic service that focuses on early diagnosis and intervention. Referrals to the team are made from primary care and a clinical diagnosis of dementia is made following a comprehensive multidisciplinary assessment including self- and informant-report HRQL using DEMQOL and DEMQOL-Proxy (Banerjee et al., 2007).

The second comprised community-dwelling older adults with dementia identified in population cohort surveys conducted by the 10/66 Dementia Research Group in the Dominican Republic, Mex-

ico, Cuba, Peru and Venezuela. Self- and informant-report HRQL were obtained in the first follow up (2007–2010) for participants identified in the baseline survey (2003–2007) to have dementia based on a battery of interview assessments (Rodriguez et al., 2008).

The study participants in both the U.K. and Latin American samples had a similar age range (see Table 1). Over half were female (63–72%). The majority (85–97%) in both study samples had mild to moderately severe dementia.

Measures

DEMQOL (28 items) and DEMQOL-Proxy (31 items) are interviewer-administered measures for obtaining self- and informant-reports on the HRQL of people with dementia. Items inquire about "feelings," "memory," and "everyday life" of the person with dementia in the last week, with four responses ranging from 1 (*a lot*) to 4 (*not at all*). Reverse scoring is required for five "positive emotion" items in DEMQOL and DEMQOL-Proxy so that higher total scores reflect better HRQL. Studies reported evidence of responsiveness, convergent validity, structural validity, and measurement reliability (Chua et al., 2016; Mulhern et al., 2013).

As part of their research program on dementia in low- and middle-income countries, the 10/66 Dementia Research Group adapted DEMQOL and DEMQOL-Proxy through forward and back-translation and administered Ibero American Spanish versions to measure HRQL in community-dwelling older adults living with dementia in the Dominican Republic, Mexico, Cuba, Peru, and Venezuela (Prina et al., 2017).

Dementia severity was summarized using the MMSE (Folstein, Folstein, & McHugh, 1975) for the U.K. sample, and the Clinical

Dementia Rating (CDR; C. P. Hughes, Berg, Danziger, Coben, & Martin, 1982) for respondents in Latin America. The MMSE is a screening tool for general cognitive impairment, with higher total scores (range = 0–30) indicating better performance, and studies have reported evidence of structural validity (Rubright, Nandakumar, & Karlawish, 2016), predictive validity, and reliability (Tombaugh & McIntyre, 1992). The CDR is a standardized semistructured interview with self- and informant inputs on cognitive and functional performance. The scale ratings (0–3) have shown evidence of criterion-validity, interrater reliability and have been validated neuropathologically for dementia (Morris, 1997). Substantial agreement between the MMSE and CDR has been documented for mild ($\kappa = 0.62$), moderate ($\kappa = 0.69$), and severe dementia ($\kappa = 0.76$; Perneczky et al., 2006).

Statistical Analysis

For cross-cultural validity, DEMQOL and DEMQOL-Proxy need to retain the same conceptual meaning, sensitivity and relevance ("measurement invariance") after they have been translated from English to Ibero American Spanish. An initial careful process of forward and backward translation was performed (Prina et al., 2017). In the present study, we tested measurement invariance by comparing HRQL data from respondents in Latin America (Ibero American Spanish) with data collected in the United Kingdom (English). In measurement invariance analyses, group comparisons are made only after matching respondents with the same HRQL estimates based on latent variable models. If group differences are found despite these matched comparisons, this suggests that U.K. and Latin American respondents differ in the way they appraise HRQL when there is no genuine difference in HRQL. As is recommended for ordered categorical data (Rhemtulla, Brosseau-

Table 1

Demographic and Clinical Characteristics of Respondents

	DEMQOL		DEMQOL-Proxy	
Variable	United Kingdom (*n* = 868)	Latin America[a] (*n* = 417)	United Kingdom (*n* = 909)	Latin America[b] (*n* = 495)
Age (*SD*)	78.6 (8.5)	79.7 (7.6)	78.9 (8.4)	79.7 (7.7)
Gender				
Male	313	126	340	138
Female	555	291	569	357
Severity				
Mild	MMSE > 20 517	CDR < 2 353	MMSE > 20 509	CDR < 2 392
Moderate	MMSE 15–20 255	CDR = 2 50	MMSE 15–20 268	CDR = 2 77
Severe	MMSE < 15 96	CDR = 3 14	MMSE < 15 132	CDR = 3 26

Note. MMSE = Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975); CDR = Clinical Dementia Rating Scale (C. P. Hughes, Berg, Danziger, Coben, & Martin, 1982). Age differences between United Kingdom and Latin America study samples have small effect sizes for DEMQOL (Hedges' *g* = .13) and DEMQOL-Proxy (Hedges' *g* = .10). Differences in proportion of male respondents between UK and Latin America study samples have small effect sizes for DEMQOL (Phi = .06) and DEMQOL-Proxy (Phi = .10). Differences in illness severity prevalence between United Kingdom and Latin America study samples have medium effect sizes for DEMQOL (Cramer's V = .25) and DEMQOL-Proxy (Cramer's V = .23).
[a] Venezuela (*n* = 53), Peru (*n* = 70), Cuba (*n* = 90), Mexico (*n* = 87), and Dominican Republic (*n* = 117). [b] Venezuela (*n* = 65), Peru (*n* = 92), Cuba (*n* = 110), Mexico (*n* = 103), and Dominican Republic (*n* = 125).

Liard, & Savalei, 2012), all latent variable models were estimated using the diagonally weighted least squares estimator with robust standard errors (denoted WLSMV in Mplus). The detail of the sequence of testing is illustrated below for DEMQOL. We provided the Mplus syntax in the online supplemental material.

Conceptual Meaning: Configural Model

The conceptual meaning of HRQL in DEMQOL item responses was first examined using single-group confirmatory factor analytic models (CFA Model 1) for the U.K. and Latin American samples separately. We chose a bifactor model framework (see Figure 1) in which all DEMQOL items load on a general factor as well as orthogonal domain factors that capture additional influence of specific item topics. As general HRQL is commonly treated as the main assessment objective in research and clinical practice (Kifley et al., 2012), the use of a bifactor model helps to retain strategic focus on general HRQL as the target construct for investigating measurement invariance (Chua et al., 2016).

Differences between these configural models provided early indications of how U.K. and Latin American respondents might differ in how they appraise HRQL. Post hoc model modifications were needed such that next stages of statistical comparisons focused on aspects of conceptual meaning that might be identical across language versions of DEMQOL. We modified these configural models based on three considerations: (a) approximate model fit (i.e., small values of root mean square error of approximation [RMSEA; <0.08] and large values of comparative fit index [CFI; >0.90]), (b) precision (i.e., standard errors) of bootstrapped model estimates (Kam & Zhou, 2016; Perera & Ganguly, 2018), and (c) scale reliability (i.e., omega-hierarchical [ω_h] coefficient; Reise, Bonifay, & Haviland, 2013). In calculating reliability for specific domain subscales, ω_h was modified as ω_s according to didactic accounts by Reise et al. (2013). However, we used the notation ω_h for both general and domain specific factor because the same methodological principle applies (Brunner, Nagy, & Wilhelm, 2012)
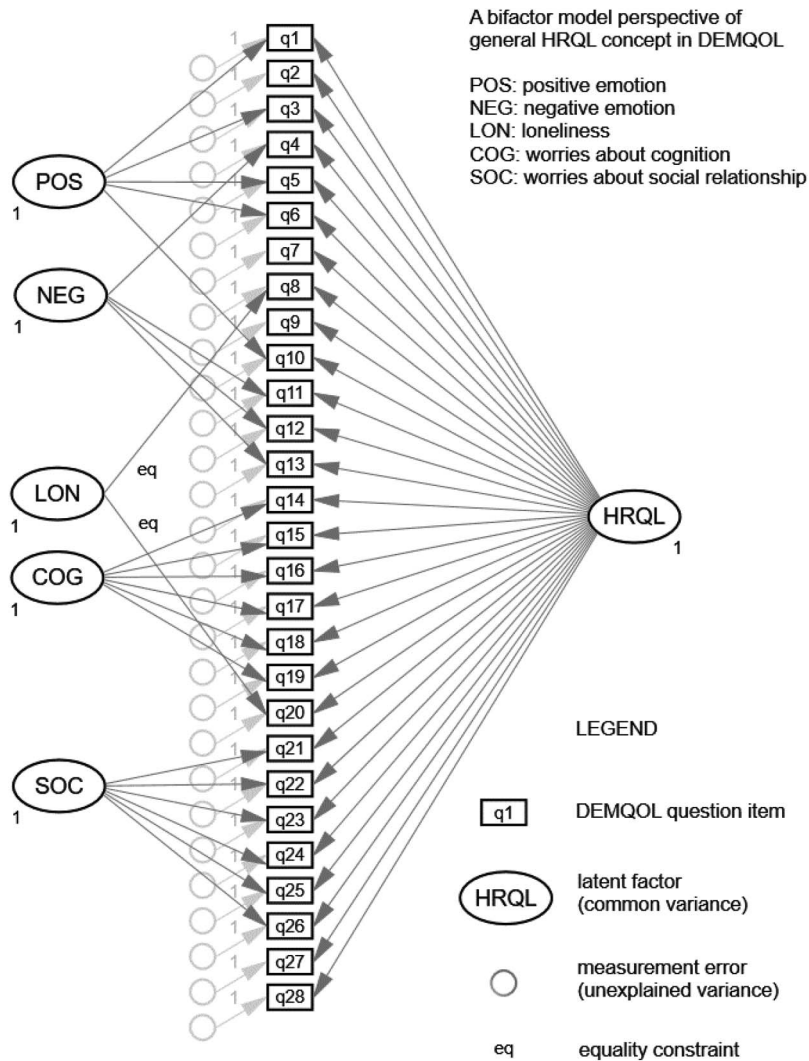

Figure 1. Bifactor model of 28-item DEMQOL (Chua et al., 2016).

and for ease of presentation. To estimate ω_h, we used standardized factor loadings estimated by WLSMV in line with recent studies across fields of psychological assessment (e.g., Adams et al., 2018; Fergus, Kelley, & Griggs, 2017; Shihata, McEvoy, & Mullan, 2018; Stanton, Forbes, & Zimmerman, 2018).

Conceptual Meaning: Configural Invariance

To see if DEMQOL items carry the same meaning for U.K. and Latin American participants, direct statistical comparisons between the configural models were made using a multiple-group CFA model. The *factor pattern configuration* refers to the underlying conceptual or cognitive frame of reference used to make item responses (Vandenberg & Lance, 2000). *Configural invariance* (Horn & McArdle, 1992) refers to an invariant pattern of factor loadings, which means that the same DEMQOL items can be grouped under identical domain themes of HRQL for the U.K. and Latin American study samples. This means that respondent groups (U.K. or Latin American) were using the same conceptual frame of reference that reflect equivalent underlying constructs (Vandenberg & Lance, 2000). If the factor loading pattern differs between United Kingdom and Latin America, the concepts that are represented by the common factors do not have the same definition (Oort, 2005). With good approximate model fit according to RMSEA and CFI values, configural invariance (Model 2a) would form the basis for saying that the conceptual meaning of HRQL is the same for both groups. Of note, a HRQL construct must carry the same meaning across groups (configural invariance) before it makes sense to examine if specific aspects are equally sensitive to individual differences in HRQL (metric invariance), and hence after also appraised according to the same internal standards (scalar invariance).

Sensitivity and Relevance: Metric and Scalar Invariance

To see if DEMQOL item scores were equally sensitive to individual differences in HRQL for the U.K. and Latin American respondents, we compared item factor loadings between the two language versions. Item factor loadings reflect the magnitude of difference in DEMQOL item scores between two individuals given their differences in HRQL. *Metric invariance* (Horn & McArdle, 1992) refers to invariant magnitude of factor loadings which means that DEMQOL items load on the same factors with the same factor loading values giving rise to identical units of measurement in the U.K. and Latin American study sample. This means that DEMQOL item scores would show the same amount of differences for U.K. or Latin American respondents given any two scenarios (e.g., poor vs. average HRQL). If the factor loading value of a DEMQOL item differs between the two study samples, then that item is more (or less) "indicative" of individual differences in HRQL due to different units of measurement between the groups (Oort, 2005). In other words, for one of the groups, the DEMQOL item is more (or less) sensitive to individual differences in HRQL. With good approximate model fit according to RMSEA and CFI values, metric invariance would form the basis for saying that the various aspects are equally sensitive to HRQL for both groups.

To see if DEMQOL items were equally relevant for U.K. and Latin American participants, we compared item thresholds between the two language versions. Item thresholds reflects item difficulty in rating "a lot"/"quite a bit"/"a little"/"not at all" for a DEMQOL item. A "difficult" item is relevant only for individuals with good HRQL. Those with poor HRQL will consistently have low item scores and thus little meaningful differences between these individuals can be observed. Conversely, an "easy" item is relevant only for individuals with poor HRQL. Those with good HRQL will consistently have high item scores and thus little meaningful differences between these individuals can be observed. *Scalar invariance* (Meredith, 1993) refers to invariant item thresholds which means that DEMQOL items load on the same factors with the same factor loading and threshold values in the U.K. and Latin American study sample. This means that internal standards for rating "a lot"/"quite a bit"/"a little"/"not at all" would be calibrated based on the same measurement origins for both groups (Oort, 2005). With good approximate model fit according to RMSEA and CFI values, scalar invariance would form the basis for saying that DEMQOL items are equally relevant for both groups because the same standards of good/poor HRQL apply (i.e., items are equally "difficult" or "easy" for both groups).

Metric and scalar invariance were tested concurrently (e.g., DuPaul et al., 2016; Fergus et al., 2017; Schroeders & Wilhelm, 2011) because item factor loadings and thresholds are mathematically interdependent (Muthén & Asparouhov, 2002; Muthén & Christoffersson, 1981). If both hypotheses were tenable, Model 2b would show good approximate fit according to RMSEA and CFI values. In addition, if Model 2b showed only trivial decline in model fit despite making more assumptions than Model 2a, this would offer stronger support for measurement invariance between versions of DEMQOL. Compared to Model 2a, the decline in exact model fit of Model 2b should not attain statistical significance based on DIFFTEST in Mplus (Asparouhov & Muthén, 2006). The decline in approximate model fit would also be considered trivial if the increases in RMSEA value was less than 0.015 and the decreases in CFI less than 0.010 (Chen, 2007; Cheung & Rensvold, 2002).

Item Characteristics

We used the combined data from U.K. and Latin American respondents to gain insights on the sensitivity and relevance of DEMQOL items. We estimated these properties with a graded-response model (Model 3), which corresponds mathematically to the CFA configural model we estimated (Kamata & Bauer, 2008). Item response theory (IRT) parameters from Model 3 were plotted to give item information curves to show (a) the level of sensitivity of a DEMQOL item (*y*-axis: discrimination parameters) and (b) how this sensitivity depends on whether people have below-average, average, and above-average HRQL (*x*-axis: difficulty parameters).

Results

Conceptual Meaning: Configural Model

All configural models showed acceptable to good fit (see Table 2), but the two samples did not show exactly the same patterns of item response. Comparing the DEMQOL single-group CFA Model 1a between the U.K. and Latin American samples (Supplemental

Table 2
Model Fit Indices

Variable	DEMQOL (English) Model χ²	RMSEA [90% CI]	CFI	DEMQOL-Proxy (English) Model χ²	RMSEA [90% CI]	CFI	DEMQOL (Spanish) Model χ²	RMSEA [90% CI]	CFI	DEMQOL-Proxy (Spanish) Model χ²	RMSEA [90% CI]	CFI
Single-group CFA												
Model 1a (df = 328)	1,420.582	.062 [.059, .065]	.918				738.266	.055 [.050, .060]	.958			
Model 1b (df = 213)	793.336	.056 [.052, .060]	.946				531.840	.060 [.054, .066]	.964			
Model 1c (df = 217)	664.715	.064 [.060, .068]	.928				568.640	.062 [.056, .069]	.961			
Multiple-group CFA[a]												
Model 2a (df = 434)	1,553.474	.063 [.060, .067]	.943									
Model 2b (df = 508)	1,788.347	.063 [.060, .066]	.935									
Single-group CFA												
Model 1a (df = 406)				1,647.018	.058 [.055, .061]	.932				1,133.244	.060 [.056, .064]	.950
Model 1b (df = 277)				715.923	.042 [.038, .046]	.972				912.963	.068 [.063, .073]	.951
Model 1c (df = 281)				1,000.566	.053 [.050, .057]	.954				944.623	.069 [.064, .074]	.948
Multiple-group CFA[b]												
Model 2a (df = 562)				1,930.082	.059 [.056, .062]	.952						
Model 2b (df = 648)				2,254.917	.059 [.057, .062]	.943						

Note. RMSEA = root mean square error of approximation; CFA = confirmatory factor analytic; CFI = comparative fit index. Model 1a: Configural model for 28-item DEMQOL and 31-item DEMQOL-Proxy. Model 1b: Configural model (without POS items) for 23-item DEMQOL and 26-item DEMQOL-Proxy. Model 1c: Configural model for 23-item DEMQOL (without NEG domain) and 26-item DEMQOL-Proxy (without social relationship [SOC] domain). Model 2a: Configural invariance. Model 2b: Configural, metric, scalar invariance.
[a] Model 2a vs. 2b: DIFFTEST χ² = 325.029 (Δ*df* = 74), *p* < .0001 for DEMQOL. [b] Model 2a vs. 2b: DIFFTEST χ² = 405.844 (Δ*df* = 86), *p* < .0001 for DEMQOL-Proxy.

Tables S1 and S2 in the online supplemental material) both had item response patterns that indicated the presence of a general HRQL factor. The U.K. sample had five additional sources of influence: positive emotion (POS), negative emotion (NEG), loneliness (LON), worries about cognition (COG), and worries about social relationship (SOC), but in the Spanish version of DEMQOL, the POS item loadings on the general HRQL factor and the NEG domain loadings were largely not statistically significant. These apparent differences between the United Kingdom and Latin America may not be major because the factor loadings are also relatively weak in the U.K. sample but might have attained statistical significance due to larger sample size.

For DEMQOL-Proxy, the same analyses showed that both groups had item response patterns that indicated the presence of a general HRQL factor (Supplemental Tables S3 and S4 in the online supplemental material). The U.K. sample had six additional sources of influence: POS, NEG, COG, SOC, and worries about finance-related tasks (FIN) and physical appearance (APP), but in the Spanish version POS item loadings on the general HRQL factor were weak and negative despite attaining statistical significance. To ensure that POS item responses were coded in the right direction for the present analyses, we repeated multiple checks and also consulted with the data owners (both for the memory clinic and 10/66 Dementia Research Group). We concluded that the coding was done correctly across the two data sets. Additional reassurance can be found in a body of literature documenting similar influences of positive and negative item wording effects on the measurement of health (Böhnke & Croudace, 2016; Lai, Garcia, Salsman, Rosenbloom, & Cella, 2012; Molina, Rodrigo, Losilla, & Vives, 2014; van Sonderen, Sanderman, & Coyne, 2013) as well as psychological traits (Marsh, 1986; Ray, Frick, Thornton, Steinberg, & Cauffman, 2016; Tomas, Oliver, Galiana, Sancho, & Lila, 2013; Weijters, Baumgartner, & Schillewaert, 2013). Of note, a recent study by an independent group of researchers used a different analytic approach (Rasch modeling) but reach similar conclusions about POS items in DEMQOL and DEMQOL-Proxy (Hendriks, Smith, Chrysanthaki, Cano, & Black, 2017). As POS items are the only reverse-worded items in DEMQOL and DEMQOL-Proxy, these model results are consistent with the presence of wording effects. Even in the U.K. sample, POS items showed the weakest factor loadings on the general HRQL factor (Chua et al., 2016).

To see if POS items were the main difference between configural models, we examined configural models (Model 1b) for the remaining 23 DEMQOL items and 26 DEMQOL-Proxy items, which resulted in good fit for both measures and study samples (see Table 2). This analysis revealed an additional difference between the two language versions. For Latin America respondents, the NEG domain factor in DEMQOL and the SOC domain factor in DEMQOL-Proxy showed signs of "factor collapse" (Chen, West, & Sousa, 2006) as indicated by weak and/or non-statistically significant loadings on the domain factors (Supplemental Tables S5–S8 in the online supplemental material). This is a statistical indication that responses to these domain items do not share any additional common variance (i.e., common theme) over and above the general theme of HRQL. Consequently, these items have sizable factor loadings only on the general HRQL factor and an additional domain factor is not retained. Even in the U.K. sample, these domain factors also had only a weak impact on item

responses as reflected by poor scale reliability ($\omega_h = 0.33$ for self-report NEG and informant-report SOC). Despite these differences in weaker sources of influence, the dominant impact of general HRQL on item responses was evident across both language versions of DEMQOL ($\omega_h = 0.87$–0.90) and DEMQOL-Proxy ($\omega_h = 0.88$–0.89).

We therefore fitted a DEMQOL bifactor model (**Model 1c**) without a NEG domain factor (Supplemental Tables S9 and S10 in the online supplemental material). For DEMQOL-Proxy, we fitted a bifactor model without a SOC domain factor (Supplemental Tables S11 and S12 in the online supplemental material). Based on Brunner et al.'s (2012) substantive interpretations of "factor collapse," we hypothesized that negative emotion is a core component of general HRQL when appraised by self-report (DEMQOL Model 1c). This is analogous to the hypothesis in cognitive psychology that 'reasoning ability' does not convey additional information (i.e., does not exist as an independent domain in a bifactor model) beyond what it conveys about individual differences in general intelligence because performance on this ability test essentially reflects only general intelligence (Gottfredson, 1997; Snow, Kyllonen, & Marshalek, 1984). In other words, responses to NEG items differ between individuals mainly because of their differences in general HRQL. Similarly, "worries about social relationship"

is a core component of general HRQL when appraised by informants (DEMQOL-Proxy Model 1c). These models showed adequate to good fit (see Table 2) and were used as the configural model for measurement invariance testing.

Conceptual Meaning: Configural Invariance

The multiple-group CFA Model 2a directly tested configural invariance between language versions of DEMQOL and DEMQOL-Proxy. The results showed good model fit (see Table 2) when we assumed the same conceptual meaning (i.e., factor loading patterns) for both language versions of DEMQOL and DEMQOL-Proxy (Supplemental Tables S13–S16 in the online supplemental material).

Sensitivity and Relevance: Metric and Scalar Invariance

When metric and scalar invariance were tested in tandem, Model 2b showed good model fit (see Table 2) for DEMQOL (see Table 3) and DEMQOL-Proxy (see Table 4). These results show it was tenable to assume that the items were equally sensitive to HRQL differences and relevant across both language versions.

Table 3

Multiple-Group CFA (Model 2b) for 23 DEMQOL Items (Unstandardized Factor Loadings With Bootstrapped Standard Errors): Scalar Invariance Estimates

Item	DEMQOL ($n = 1,284$), EL ($n = 867$), ES ($n = 417$)	h^2	GEN	*SE*	COG	*SE*	LON	*SE*	SOC	*SE*
2	Worried or anxious	.38	.97	.05						
4	Frustrated	.41	.89	.04						
7	Sad	.39	.95	.04						
8	Lonely	.59	.73	.05			1.00	a		
9	Distressed	.37	1.12	.05						
11	Irritable	.42	.90	.04						
12	Fed-up	.49	1.00	a						
13	Things to do but couldn't	.46	.70	.05						
14	Forget recent things	.69	.85	.05	1.07	.11				
15	Forgetting who people are	.78	.80	.05	.96	.09				
16	Forgetting what day it is	.70	.72	.05	1.00	a				
17	Your thoughts being muddled	.79	.95	.05	1.03	.09				
18	Difficulty making decisions	.71	1.01	.05	.79	.10				
19	Poor concentration	.74	.95	.05	.82	.11				
20	Not having enough company	.77	.74	.05			1.00	a		
21	Get on with people close	.70	.90	.07					.70	.13
22	Getting affection that you want	.68	.89	.07					.95	.16
23	People not listening to you	.87	.86	.06					1.00	a
24	Making yourself understood	.82	.85	.06					.77	.07
25	Getting help when you need it	.73	.99	.06					.77	.09
26	Getting to the toilet in time	.50	.72	.07					.61	.11
27	How you feel in yourself	.59	1.08	.05						
28	Your health overall	.44	.91	.05						
	Factor variance (EL)		.48		.23		.53		.32	
	Factor variance (ES)		1.52		.88		.43		.48	
	Factor means (ES)		.35		.16ns		−.48		−.51	

Note. GEN = general HRQL; COG = worries about cognition; LON = loneliness; SOC = worries about social relationship. Model fit from non-bootstrapped results: $\chi^2 = 1{,}788.347$ ($df = 508$), English (EL) $\chi^2 = 983.708$, Spanish (ES) $\chi^2 = 804.639$. Root mean square error of approximation (RMSEA) = .063 (90% confidence interval [CI] = .060, .066), comparative fit index (CFI) = .935. DIFFTEST $\chi^2 = 325.029$ ($\Delta df = 74$), $p < .0001$. h^2 = communalities.
[a] Unstandardized factor loading fixed at value of 1. Numbers in italics are values of Standard Error (*SE*).

Table 4

Multiple-Group CFA (Model 2b) for 26 DEMQOL-Proxy Items (Unstandardised Factor Loadings With Bootstrapped Standard Errors): Scalar Invariance Estimates

Item	DEMQOL ($n = 1404$), DEMQOL-Proxy EL ($n = 909$), DEMQOL-Proxy ES ($n = 495$)	h^2	GEN	*SE*	NEG	*SE*	APP	*SE*	FIN	*SE*	COG	*SE*
2	Worried or anxious	.57	.73	*.05*	1.00	[a]						
3	Frustrated	.56	.64	*.05*	1.21	*.11*						
5	Sad	.52	.71	*.06*	1.20	*.10*						
7	Distressed	.43	.84	*.06*	1.01	*.08*						
9	Irritable	.26	.47	*.06*	1.07	*.12*						
10	Fed-up	.27	.80	*.05*	1.06	*.10*						
12	Memory in general	.60	.76	*.06*							1.00	[a]
13	Forget long ago things	.71	.68	*.06*							.79	*.08*
14	Forget recent things	.78	.88	*.06*							1.21	*.07*
15	Forget people's names	.80	.77	*.05*							.81	*.08*
16	Forget where he/she is	.73	.77	*.06*							.60	*.08*
17	Forget what day it is	.68	.87	*.05*							.75	*.08*
18	Thoughts muddled	.71	.98	*.05*							.81	*.10*
19	Difficulty deciding	.68	.94	*.05*							.76	*.10*
20	Making self understood	.58	.87	*.05*							.57	*.10*
21	Keeping clean	.89	.76	*.07*			1.00	[a]				
22	Keeping looking nice	.94	.78	*.06*			1.00	[a]				
23	Get things from shops	.72	.88	*.05*					1.00	[a]		
24	Using money to pay	.88	.88	*.06*					1.31	*.12*		
25	Looking after finances	.85	.83	*.06*					1.31	*.12*		
26	Things take longer	.49	.95	*.05*								
27	Get in touch with people	.57	1.00	[a]								
28	Not enough company	.40	.88	*.05*								
29	Not being able to help	.69	.91	*.04*								
30	Not playing a useful part	.68	.92	*.05*								
31	His/her physical health	.53	.72	*.05*								
	Factor variance (EL)		.52		.25		.51		.21		.28	
	Factor variance (ES)		1.80		1.15		1.66		1.04		2.43	
	Factor means (ES)		.12[ns]		.84		−.87		.20[ns]		1.13	

Note. GEN = general HRQL; NEG = negative emotion; APP = worries about physical appearance; FIN = worries about finance-related tasks; COG = worries about cognition. Model fit from non-bootstrapped results: $\chi^2 = 2{,}254.917$ ($df = 648$), English (EL) $\chi^2 = 934.728$, Spanish (ES) $\chi^2 = 1{,}320.188$. Root mean square error of approximation (RMSEA) = .059 (90% confidence interval [CI] = .057, .062), comparative fit index (CFI) = .943. DIFFTEST $\chi^2 = 405.844$ ($\Delta df = 86$), $p < .0001$. h^2 = communalities.
[a] Unstandardized factor loading fixed at value of 1. Numbers in italics are values of Standard Error (*SE*).

Compared to Model 2a, the decline in exact model fit of Model 2b according to the DIFFTEST attained statistical significance (see Table 2). However, the decline in approximate model fit was considered trivial for DEMQOL (RMSEA: 0.063 vs. 0.063 and CFI: 0.943 vs. 0.935) and DEMQOL-Proxy (RMSEA: 0.059 vs. 0.059 and CFI: 0.952 vs. 0.943; Chen, 2007; Cheung & Rensvold, 2002). Taken together, these criteria lend support for the tenability of measurement invariance (Model 2b).

Item Characteristics

With tenable support for measurement invariance in a subset of DEMQOL and DEMQOL-Proxy items, we used the combined U.K. and Latin American data to estimate an IRT graded response model (Model 3). In this model linguistic group was treated as an external covariate predicting differences in latent means between U.K. and Latin American respondents. IRT parameters from Model 3 (Supplemental Tables S17 and S18 in the online supplemental material) were plotted to give item information curves.

Figures 2 and 3 show the item information curves for "worries about cognition" items in DEMQOL and DEMQOL-Proxy items respectively. Across the *x*-axis, latent model estimates of HRQL (general factor; Model 3) are standardized so that sample average is located at the mean of 0 with a standard deviation of 1. The sensitivity (*y*-axis) of most DEMQOL and DEMQOL-Proxy items rises to the highest level at around 1 *SD* below the sample average HRQL (*x*-axis). This means DEMQOL and DEMQOL-Proxy measurements were most sensitive for detecting HRQL differences between people with below average HRQL.

Among 23 DEMQOL items, responses about "worries about cognition" (see Figure 2) as well as "negative emotion" (Supplemental Figure S2A in the online supplemental material) were more sensitive for detecting HRQL differences between people with above average HRQL (i.e., 1 *SD* above sample average). Items for "worries about social relationship" (Supplemental Figure S2B in the online supplemental material) were particularly useful for detecting HRQL differences between people with below average HRQL. However, Item 26 (getting to toilet in time) was easy even for respondents with significant HRQL impairment to report "not at all" worried about this matter, so this item was mainly relevant
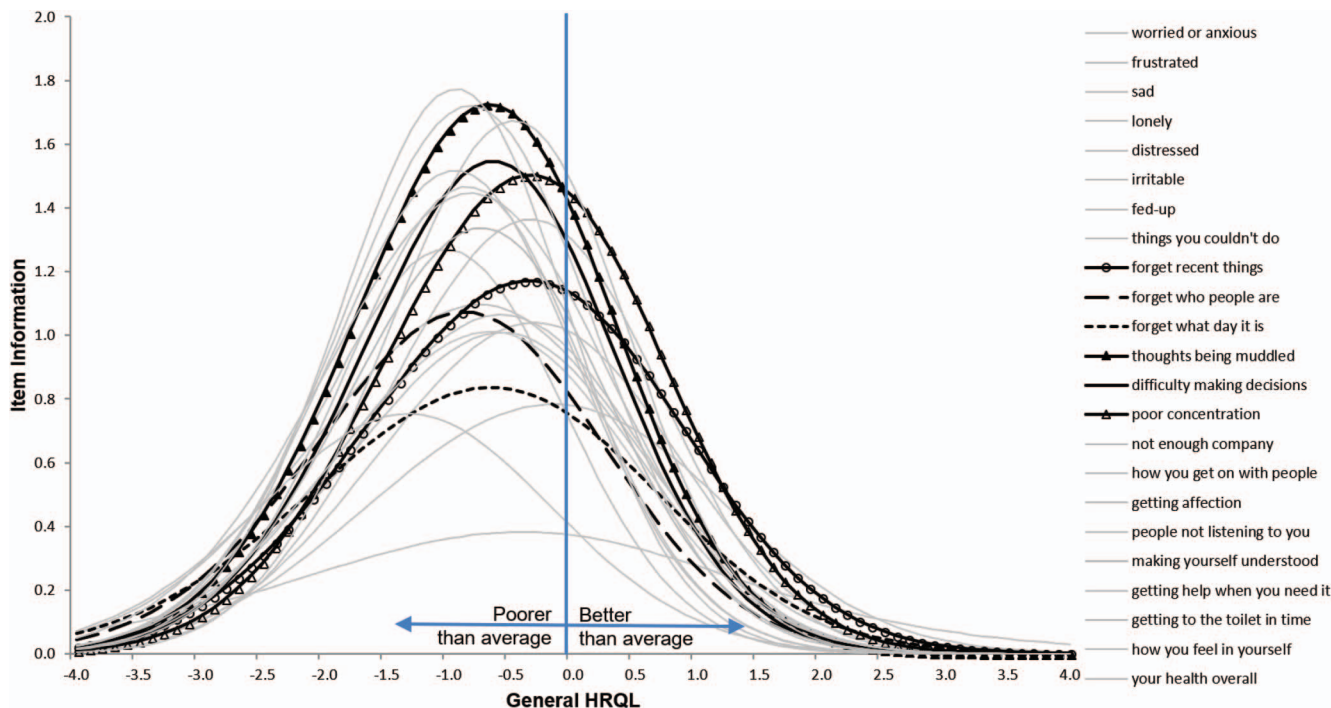
Figure 2. Item information curves of six "worries about cognition" items for self-report HRQL (DEMQOL Model 3). HRQL = health related quality of life. See the online article for the color version of this figure.

for severe impairment (over 1 *SD* below average). Most other items (Supplemental Figure S2C and S2D in the online supplemental material) show similar levels of sensitivity and standards of difficulty.

Among 26 DEMQOL-Proxy items, informant ratings of "worries about cognition" (see Figure 3) and "worries about social relationship" (Supplemental Figure S3a in the online supplemental material) were more sensitive for detecting HRQL differences between people with above average HRQL. "Negative emotion" items show the least sensitivity across HRQL levels (Figure S3b in online supplement material). Previous research suggested that affective states are less easily observed by informants (Novella et al., 2001). For worries about "finance-related tasks" and "physical appearance" (Supplemental Figure S3C–S3D in the online supplemental material) most informants would rate "not at all" on these items, so they are mainly relevant for assessing severe impairment (over 1 *SD* below average).

Discussion

The main finding of this study is that the data offer the first empirical support for the use of a dementia-specific measure of HRQL cross-culturally, in this case the use of the DEMQOL system in the United Kingdom and Latin America. This is supported by our psychometric evaluation which found strong measurement invariance for the general HRQL factor, the dominant influence on item responses from self- and informant-reports. We can therefore conclude that DEMQOL and DEMQOL-Proxy carry the same meaning, sensitivity, and rel-

evance for respondents in the United Kingdom and Latin America. However, differences in domain factors suggest benefits in making statistical adjustments for weaker influences on item responses. Also, these cross-cultural comparisons provide new insights on HRQL measurement in dementia, showing that "negative emotion" is a core component in self-reports and "worries about social relationship" in informant-reports.

Lawton (1994) postulated that the absence of HRQL impairment is not the same as good HRQL and that HRQL is a construct "concerned primarily with decrements from the average." Our construct validation study supports Lawton's position. We found that "positive emotion" was not a major component of general HRQL in the Latin American sample, a similar pattern to that found in U.K. samples (Chua et al., 2016). If the absence of "HRQL impairment" is not the same as "good HRQL," then HRQL impairment might be considered a unipolar construct (Reise & Waller, 2009) in which the presence of impairment shows meaningful individual differences, but the absence of impairment gives little insight about what constitutes "good" HRQL. This is in line with the finding in the general HRQL measurement literature that negative and positive components of well-being may be different or partly independent aspects of people's experience (Böhnke & Croudace, 2016).

A further explanation for the factor loadings may be that there are wording effects for the "positive emotion" because they are reverse-worded items in DEMQOL and DEMQOL-Proxy. In the absence of HRQL impairment (i.e., "good" HRQL), one may find it easy to respond "not at all" when asked
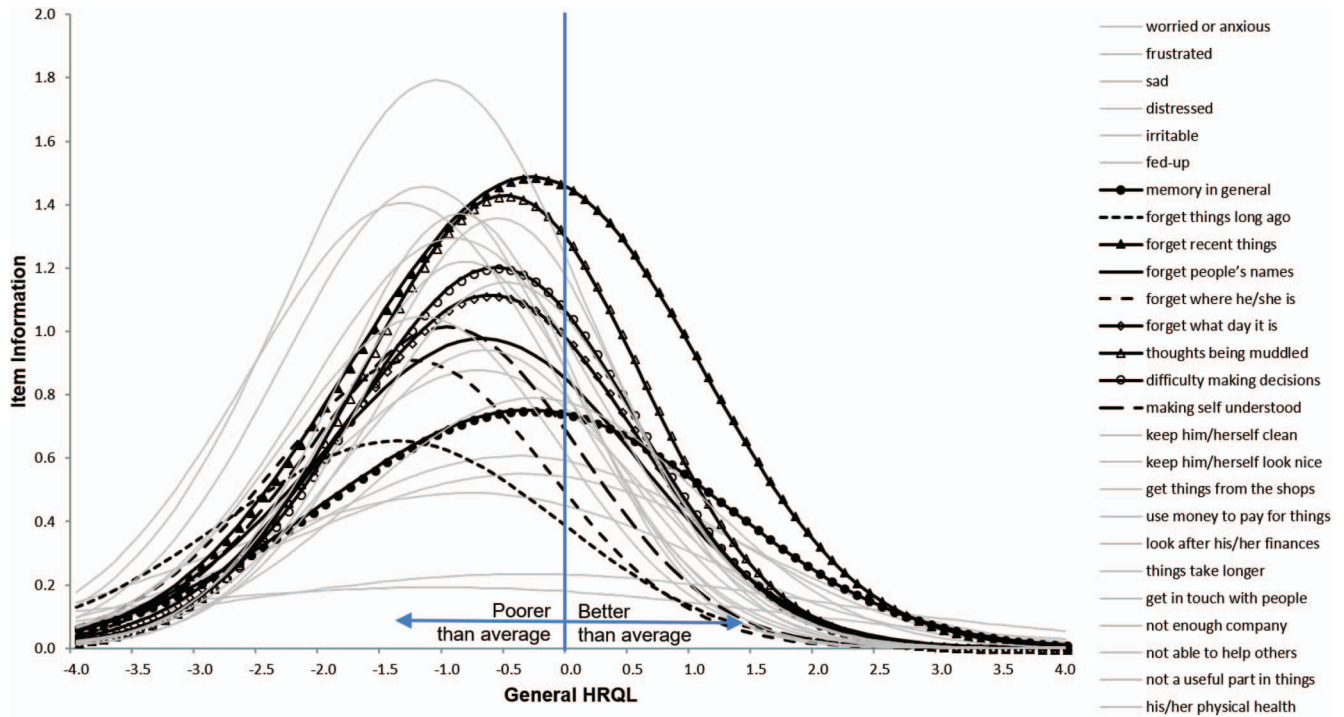
Figure 3. Item information curves of nine "worries about cognition" items for informant-report HRQL (DEMQOL-Proxy Model 3). HRQL = health related quality of life. See the online article for the color version of this figure.

if he or she has "worries" but not as easy to respond "very much" when asked if he or she is "feeling cheerful". This is consistent with findings from a U.K. population-based study which showed an asymmetry between strong adverse reactions to deteriorations in health, alongside weak increases in well-being after health improvements (Binder & Coad, 2013). Such wording effects may have unequal strengths in different languages. Consideration should be given to these issues in the development and translation of instruments for cross-cultural use.

Further discussion, informed by evidence, is needed before POS items can be recommended for exclusion from the questionnaire. Our findings do not mean that positive states are not important for general HRQL. Lawton (1994) proposed that, for people with dementia, indicators of positive states may be found in both positive affective states and positive behaviors, such as behaviors that exemplify social engagement. As such DEMQOL and DEMQOL-Proxy consider positive states as part of general HRQL by tapping on items that focus on "worries about social relationship." This focus contributes to the clinical relevance of HRQL assessment as social functioning is "a treatment goal that seems appropriate for an illness whose manifestations in general appear to represent estrangement from the external world" (Lawton, 1994). Such a focus is also consistent with a large body of literature demonstrating that social functioning plays a pivotal role in the illness experience (Frick, Irving, & Rehm, 2012; Lou, Chi, Kwan, & Leung, 2013; MacRae, 2011; T. F. Hughes, Flatt, Fu, Chang, & Ganguli, 2013) as well as healthy aging in general (Coyle & Dugan,

2012; Huxhold, Fiori, & Windsor, 2013; Ichida et al., 2013; Rook, Luong, Sorkin, Newsom, & Krause, 2012).

Study Limitations

This study has three important limitations. First, although the high overall sample size for the Latin American countries was appropriate for invariance analyses ($n = 417$ for DEMQOL and $n = 495$ for DEMQOL-Proxy), the numbers in individual country samples were relatively small (between $n = 56$ for DEMQOL in Venezuela and $n = 125$ for DEMQOL-Proxy in the Dominican Republic). We therefore carried out pooled analyses for the Latin American sample on the basis that the same translation was used, however this means we cannot comment on between-country differences. Second, the samples were recruited using different processes in the U.K. and Latin American sites. The former was from a memory assessment service (Banerjee et al., 2007) and the latter from a program of population research (Prince et al., 2007). However, all had well-characterized diagnoses of dementia and statistically matched comparisons were used. Third, although our sequence of models was based on established strategies to test for measurement invariance (Vandenberg & Lance, 2000), modeling decisions were data-driven and need replication in independent samples to guard against sample-based overfitting (Borsboom, 2006). Of note, future studies should consider a priori use of bifactor (S-1) models (Eid, Geiser, Koch, & Heene, 2017) to help clarify if negative emotion and/or worries about social functioning constitute the core meaning of HRQL in dementia

(Heinrich, Zagorscak, Eid, & Knaevelsrud, 2018). Neverthe-less, the study also has strengths. We assembled a unique dataset which allowed for the novel investigation of a key concern in global dementia research (Dichter et al., 2016). Our findings align with empirical literature that shows that even for well-developed measures with translation processes that follow best-practice guidelines, international comparability is not a straightforward issue (Romppel et al., 2017; Stevanovic et al., 2017; Yao et al., 2018) but that there is room for optimism that HRQL measures can be used cross-culturally in dementia.

Conclusions

Treatment and policy interventions that improve the lives of people with dementia carry both societal and fiscal impact. The stakes are particularly high in world regions like Latin America where the global burden of dementia is high and growing quickly (Prince, Wimo, et al., 2015). To develop global strat-egies, HRQL assessment is therefore as needed in low- and middle-income countries as it is in high-income countries. The lack of research resources in low- and middle-income regions like Latin America (Barreto et al., 2012) is a key challenge to developing an evidence base on interventions in dementia that is relevant to the countries in which they may be deployed. This study presents the first in-depth study of the cross-cultural assessment of HRQL and shows that, with care, using translated instruments can generate meaningful insights. This is an im-portant step on the path to developing a firm empirical basis for the benefits of dementia interventions in low- and middle-income countries as well as future global trials.

References

Adams, L. M., Wilson, T. E., Merenstein, D., Milam, J., Cohen, J., Golub, E. T., . . . Cook, J. A. (2018). Using the Center for Epidemiologic Studies Depression Scale to assess depression in women with HIV and women at risk for HIV: Are somatic items invariant? *Psychological Assessment, 30,* 97–105. http://dx.doi.org/10.1037/pas0000456

The Alzheimer's Association. (2013). 2013 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Asso-ciation, 9,* 208–245. http://dx.doi.org/10.1016/j.jalz.2013.02.003

Asparouhov, T., & Muthén, B. (2006). *Robust chi square difference testing with mean and variance adjusted test statistics* (Mplus Web Notes: No. 10). Retrieved from http://www.statmodel.com/download/webnotes/webnote10.pdf

Ballard, C., Hanney, M. L., Theodoulou, M., Douglas, S., McShane, R., Kossakowski, K., . . . the DART-AD investigators. (2009). The dementia antipsychotic withdrawal trial (DART-AD): Long-term follow-up of a randomised placebo-controlled trial. *The Lancet Neurology, 8,* 151–157. http://dx.doi.org/10.1016/S1474-4422(08)70295-3

Banerjee, S., Hellier, J., Dewey, M., Romeo, R., Ballard, C., Baldwin, R., . . . Burns, A. (2011). Sertraline or mirtazapine for depression in dementia (HTA-SADD): A randomised, multicentre, double-blind, placebo-controlled trial. *The Lancet, 378,* 403–411. http://dx.doi.org/10.1016/S0140-6736(11)60830-1

Banerjee, S., Willis, R., Matthews, D., Contell, F., Chan, J., & Murray, J. (2007). Improving the quality of care for mild to moderate dementia: An evaluation of the Croydon Memory Service Model. *International Jour-nal of Geriatric Psychiatry, 22,* 782–788. http://dx.doi.org/10.1002/gps .1741

Barreto, S. M., Miranda, J. J., Figueroa, J. P., Schmidt, M. I., Munoz, S., Kuri-Morales, P. P., & Silva, J. B., Jr. (2012). Epidemiology in Latin

America and the Caribbean: Current situation and challenges. *Interna-tional Journal of Epidemiology, 41,* 557–571. http://dx.doi.org/10.1093/ije/dys017

Bentvelzen, A., Aerts, L., Seeher, K., Wesson, J., & Brodaty, H. (2017). A comprehensive review of the quality and feasibility of dementia assess-ment measures: The Dementia Outcomes Measurement Suite. *Journal of the American Medical Directors Association, 18,* 826–837. http://dx.doi .org/10.1016/j.jamda.2017.01.006

Binder, M., & Coad, A. (2013). "I'm afraid I have bad news for you . . .:" Estimating the impact of different health impairments on subjective well-being. *Social Science & Medicine, 87,* 155–167. http://dx.doi.org/10.1016/j.socscimed.2013.03.025

Böhnke, J. R., & Croudace, T. J. (2016). Calibrating well-being, quality of life and common mental disorder items: Psychometric epidemiology in public mental health research. *The British Journal of Psychiatry, 209,* 162–168. http://dx.doi.org/10.1192/bjp.bp.115.165530

Borsboom, D. (2006). When does measurement invariance matter? *Medi-cal Care, 44*(11, Suppl. 3), S176–S181. http://dx.doi.org/10.1097/01.mlr .0000245143.08679.cc

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80,* 796–846. http://dx.doi .org/10.1111/j.1467-6494.2011.00749.x

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504. http://dx.doi.org/10.1080/10705510701301834

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41,* 189–225. http://dx.doi.org/10.1207/s15327906 mbr4102_5

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Mod-eling, 9,* 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Chua, K. C., Brown, A., Little, R., Matthews, D., Morton, L., Loftus, V., . . . Banerjee, S. (2016). Quality-of-life assessment in dementia: The use of DEMQOL and DEMQOL-Proxy total scores. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treat-ment, Care and Rehabilitation, 25,* 3107–3118. http://dx.doi.org/10 .1007/s11136-016-1343-1

Clark, P. G. (1995). Quality of life, values, and teamwork in geriatric care: Do we communicate what we mean? *The Gerontologist, 35,* 402–411. http://dx.doi.org/10.1093/geront/35.3.402

Cooper, C., Mukadam, N., Katona, C., Lyketsos, C. G., Ames, D., Rabins, P., . . . the World Federation of Biological Psychiatry–Old Age Task-force. (2012). Systematic review of the effectiveness of non-pharmacological interventions to improve quality of life of people with dementia. *International Psychogeriatrics, 24,* 856–870. http://dx.doi .org/10.1017/S1041610211002614

Coyle, C. E., & Dugan, E. (2012). Social isolation, loneliness and health among older adults. *Journal of Aging and Health, 24,* 1346–1363. http://dx.doi.org/10.1177/0898264312460275

Dai, T., Davey, A., Woodard, J. L., Miller, L. S., Gondo, Y., Kim, S. H., & Poon, L. W. (2013). Sources of variation on the mini-mental state examination in a population-based sample of centenarians. *Journal of the American Geriatrics Society, 61,* 1369–1376. http://dx.doi.org/10 .1111/jgs.12370

Dichter, M. N., Schwab, C. G., Meyer, G., Bartholomeyczik, S., & Halek, M. (2016). Linguistic validation and reliability properties are weak investigated of most dementia-specific quality of life measurements-a systematic review. *Journal of Clinical Epidemiology, 70,* 233–245. http://dx.doi.org/10.1016/j.jclinepi.2015.08.002

DuPaul, G. J., Reid, R., Anastopoulos, A. D., Lambert, M. C., Watkins, M. W., & Power, T. J. (2016). Parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms: Factor structure and normative

data. *Psychological Assessment, 28,* 214–225. http://dx.doi.org/10.1037/pas0000166

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22,* 541–562. http://dx.doi.org/10.1037/met0000083

Fergus, T. A., Kelley, L. P., & Griggs, J. O. (2017). Examining the ethnoracial invariance of a bifactor model of anxiety sensitivity and the incremental validity of the physical domain-specific factor in a primary-care patient sample. *Psychological Assessment, 29,* 1290–1295. http://dx.doi.org/10.1037/pas0000431

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

Frick, U., Irving, H., & Rehm, J. (2012). Social relationships as a major determinant in the valuation of health states. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 21,* 209–213. http://dx.doi.org/10.1007/s11136-011-9945-0

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24,* 13–23. http://dx.doi.org/10.1016/S0160-2896(97)90011-8

Harrison, J. K., Noel-Storr, A. H., Demeyere, N., Reynish, E. L., & Quinn, T. J. (2016). Outcomes measures in a decade of dementia and mild cognitive impairment trials. *Alzheimer's Research & Therapy, 8,* 48. http://dx.doi.org/10.1186/s13195-016-0216-8

Heinrich, M., Zagorscak, P., Eid, M., & Knaevelsrud, C. (2018). Giving G a meaning: An application of the bifactor-(S-1) approach to realize a more symptom-oriented modeling of the Beck Depression Inventory-II. *Assessment*. Advance online publication. http://dx.doi.org/10.1177/1073191118803738

Hendriks, A. A. J., Smith, S. C., Chrysanthaki, T., Cano, S. J., & Black, N. (2017). DEMQOL and DEMQOL-Proxy: A Rasch analysis. *Health and Quality of Life Outcomes, 15,* 164. http://dx.doi.org/10.1186/s12955-017-0733-6

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18,* 117–144. http://dx.doi.org/10.1080/03610739208253916

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *The British Journal of Psychiatry, 140,* 566–572. http://dx.doi.org/10.1192/bjp.140.6.566

Hughes, T. F., Flatt, J. D., Fu, B., Chang, C. C., & Ganguli, M. (2013). Engagement in social activities and progression from mild to severe cognitive impairment: The MYHAT study. *International Psychogeriatrics, 25,* 587–595. http://dx.doi.org/10.1017/S1041610212002086

Huxhold, O., Fiori, K. L., & Windsor, T. D. (2013). The dynamic interplay of social network characteristics, subjective well-being, and health: The costs and benefits of socio-emotional selectivity. *Psychology and Aging, 28,* 3–16. http://dx.doi.org/10.1037/a0030170

Ichida, Y., Hirai, H., Kondo, K., Kawachi, I., Takeda, T., & Endo, H. (2013). Does social participation improve self-rated health in the older population? A quasi-experimental intervention study. *Social Science & Medicine, 94,* 83–90. http://dx.doi.org/10.1016/j.socscimed.2013.05.006

Jones, R. N. (2006). Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination: Detecting differential item functioning using MIMIC modeling. *Medical Care, 44*(11, Suppl. 3), S124–S133. http://dx.doi.org/10.1097/01.mlr.0000245250.50114.0f

Kam, C. C., & Zhou, M. (2016). Is the dark triad better studied using a variable- or a person-centered approach? An exploratory investigation. *PLoS ONE, 11*(8), e0161628. http://dx.doi.org/10.1371/journal.pone.0161628

Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling, 15,* 136–153. http://dx.doi.org/10.1080/10705510701758406

Kifley, A., Heller, G. Z., Beath, K. J., Bulger, D., Ma, J., & Gebski, V. (2012). Multilevel latent variable models for global health-related quality of life assessment. *Statistics in Medicine, 31,* 1249–1264. http://dx.doi.org/10.1002/sim.4455

Knapp, M., Iemmi, V., & Romeo, R. (2013). Dementia care costs and outcomes: A systematic review. *International Journal of Geriatric Psychiatry, 28,* 551–561. http://dx.doi.org/10.1002/gps.3864

Knapp, M., Prince, M., Albanese, E., Banerjee, S., Dhanasiri, S., Fernandez, J.-L., . . . Stewart, R. (2007). *Dementia U. K. Alzheimer's Society*. Retrieved from http://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2

Lai, J. S., Garcia, S. F., Salsman, J. M., Rosenbloom, S., & Cella, D. (2012). The psychosocial impact of cancer: Evidence in support of independent general positive and negative components. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 21,* 195–207. http://dx.doi.org/10.1007/s11136-011-9935-2

Lawton, M. P. (1994). Quality of life in Alzheimer disease. *Alzheimer disease and associated disorders, 8*(Suppl. 3), 138–150.

Lou, V. W., Chi, I., Kwan, C. W., & Leung, A. Y. (2013). Trajectories of social engagement and depressive symptoms among long-term care facility residents in Hong Kong. *Age and Ageing, 42,* 215–222. http://dx.doi.org/10.1093/ageing/afs159

MacRae, H. (2011). Self and other: The importance of social interaction and social relationships in shaping the experience of early-stage Alzheimer's disease. *Journal of Aging Studies, 25,* 445–456. http://dx.doi.org/10.1016/j.jaging.2011.06.001

Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive developmental phenomenon. *Developmental Psychology, 22,* 37–49. http://dx.doi.org/10.1037/0012-1649.22.1.37

Meredith, W. (1993). Measurement invariance, factor-analysis and factorial invariance. *Psychometrika, 58,* 525–543. http://dx.doi.org/10.1007/BF02294825

Missotten, P., Dupuis, G., & Adam, S. (2016). Dementia-specific quality of life instruments: A conceptual analysis. *International Psychogeriatrics, 28,* 1245–1262. http://dx.doi.org/10.1017/S1041610216000417

Molina, J. G., Rodrigo, M. F., Losilla, J. M., & Vives, J. (2014). Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychological Assessment, 26,* 1031–1037. http://dx.doi.org/10.1037/a0036472

Morris, J. C. (1997). Clinical dementia rating: A reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International Psychogeriatrics, 9*(Suppl. 1), 173–176. http://dx.doi.org/10.1017/s1041610297004870

Mulhern, B., Rowen, D., Brazier, J., Smith, S., Romeo, R., Tait, R., . . . Banerjee, S. (2013). Development of DEMQOL-U and DEMQOL-PROXY-U: Generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. *Health Technology Assessment, 17,* 1–140. http://dx.doi.org/10.3310/hta17050

Muthén, B., & Asparouhov, T. (2002). *Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus* (Mplus Web Notes: No. 4). Retrieved from http://www.statmodel.com/examples/webnote.shtml

Muthén, B., & Christoffersson, A. (1981). Simultaneous factor-analysis of dichotomous-variables in several groups. *Psychometrika, 46,* 407–419. http://dx.doi.org/10.1007/BF02293798

Novella, J. L., Jochum, C., Jolly, D., Morrone, I., Ankri, J., Bureau, F., & Blanchard, F. (2001). Agreement between patients' and proxies' reports of quality of life in Alzheimer's disease. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care &*

Rehabilitation, 10, 443–452. http://dx.doi.org/10.1023/A:1012522013817

Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 14,* 587–598. http://dx.doi.org/10.1007/s11136-004-0830-y

Perera, H. N., & Ganguly, R. (2018). Construct validity of scores from the connor-davidson resilience scale in a sample of postsecondary students with disabilities. *Assessment, 25,* 193–205. http://dx.doi.org/10.1177/1073191116646444

Perneczky, R., Wagenpfeil, S., Komossa, K., Grimmer, T., Diehl, J., & Kurz, A. (2006). Mapping scores onto stages: Mini-mental state examination and clinical dementia rating. *The American Journal of Geriatric Psychiatry, 14,* 139–144. http://dx.doi.org/10.1097/01.JGP.0000192478.82189.a8

Perng, C. H., Chang, Y. C., & Tzang, R. F. (2018). The treatment of cognitive dysfunction in dementia: A multiple treatments meta-analysis. *Psychopharmacology, 235,* 1571–1580. http://dx.doi.org/10.1007/s00213-018-4867-y

Prina, A. M., Acosta, D., Acosta, I., Guerra, M., Huang, Y., Jotheeswaran, A. T., . . . Prince, M. (2017). Cohort profile: The 10/66 study. *International Journal of Epidemiology, 46,* 406–406i.

Prince, M. (2008). Measurement validity in cross-cultural comparative research. *Epidemiologia e Psichiatria Sociale, 17,* 211–220. http://dx.doi.org/10.1017/S1121189X00001305

Prince, M., Ferri, C. P., Acosta, D., Albanese, E., Arizaga, R., Dewey, M., . . . Uwakwe, R. (2007). The protocols for the 10/66 Dementia Research Group population-based research programme. *BMC Public Health, 7,* 165. http://dx.doi.org/10.1186/1471-2458-7-165

Prince, M., Guerchet, M., & Prina, A. M. (2015, March). The epidemiology and impact of dementia: Current state and future trends. *Thematic Briefs for the First WHO Ministerial Conference on Global Action Against Dementia.* Advance online publication. Retrieved from http://www.who.int/mental_health/neurology/dementia/thematic_briefs_dementia/en/

Prince, M., Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T., & Prina, A. M. (2015). *The World Alzheimer Report 2015, the global impact of dementia: An analysis of prevalence, incidence, cost and trends.* Retrieved from https://www.alz.co.uk/research/world-report-2015

Ray, J. V., Frick, P. J., Thornton, L. C., Steinberg, L., & Cauffman, E. (2016). Positive and negative item wording and its influence on the assessment of callous-unemotional traits. *Psychological Assessment, 28,* 394–404. http://dx.doi.org/10.1037/pas0000183

Reilly, S., Harding, A., Morbey, H., Ahmed, F., Leroi, I., Williamson, P., . . . Challis, D. (2014). The development of a Dementia core outcome set for dementia care in the community. *COMET Initiative.* Retrieved from http://www.comet-initiative.org/studies/details/677

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. http://dx.doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95,* 129–140. http://dx.doi.org/10.1080/00223891.2012.725437

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5,* 27–48. http://dx.doi.org/10.1146/annurev.clinpsy.032408.153553

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17,* 354–373. http://dx.doi.org/10.1037/a0029315

Rodriguez, J. J. l., Ferri, C. P., Acosta, D., Guerra, M., Huang, Y., Jacob, K. S., . . . the 10/66 Dementia Research Group. (2008). Prevalence of dementia in Latin America, India, and China: A population-based cross-sectional survey. *The Lancet, 372,* 464–474. http://dx.doi.org/10.1016/S0140-6736(08)61002-8

Romppel, M., Hinz, A., Finck, C., Young, J., Brähler, E., & Glaesmer, H. (2017). Cross-cultural measurement invariance of the General Health Questionnaire-12 in a German and a Colombian population sample. *International Journal of Methods in Psychiatric Research, 26*(4), e1532. http://dx.doi.org/10.1002/mpr.1532

Rook, K. S., Luong, G., Sorkin, D. H., Newsom, J. T., & Krause, N. (2012). Ambivalent versus problematic social ties: Implications for psychological health, functional health, and interpersonal coping. *Psychology and Aging, 27,* 912–923. http://dx.doi.org/10.1037/a0029246

Rubright, J. D., Nandakumar, R., & Karlawish, J. (2016). Identifying an appropriate measurement modeling approach for the Mini-Mental State Examination. *Psychological Assessment, 28,* 125–133. http://dx.doi.org/10.1037/pas0000146

Schroeders, U., & Wilhelm, O. (2011). Equivalence of reading and listening comprehension across test media. *Educational and Psychological Measurement, 71,* 849–869. http://dx.doi.org/10.1177/0013164410391468

Shihata, S., McEvoy, P. M., & Mullan, B. A. (2018). A bifactor model of intolerance of uncertainty in undergraduate and clinical samples: Do we need to reconsider the two-factor model? *Psychological Assessment, 30,* 893–903. http://dx.doi.org/10.1037/pas0000540

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 47–103). Hillsdale, NJ: Erlbaum.

Spijker, A., Vernooij-Dassen, M., Vasse, E., Adang, E., Wollersheim, H., Grol, R., & Verhey, F. (2008). Effectiveness of nonpharmacological interventions in delaying the institutionalization of patients with dementia: A meta-analysis. *Journal of the American Geriatrics Society, 56,* 1116–1128. http://dx.doi.org/10.1111/j.1532-5415.2008.01705.x

Stanton, K., Forbes, M. K., & Zimmerman, M. (2018). Distinct dimensions defining the Adult ADHD Self-Report Scale: Implications for assessing inattentive and hyperactive/impulsive symptoms. *Psychological Assessment, 30,* 1549–1559. http://dx.doi.org/10.1037/pas0000604

Stevanovic, D., Bagheri, Z., Atilola, O., Vostanis, P., Stupar, D., Moreira, P., . . . Ribas, R. (2017). Cross-cultural measurement invariance of the Revised Child Anxiety and Depression Scale across 11 world-wide societies. *Epidemiology and Psychiatric Sciences, 26,* 430–440. http://dx.doi.org/10.1017/S204579601600038X

Tomas, J. M., Oliver, A., Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling, 20,* 299–313. http://dx.doi.org/10.1080/10705511.2013.769394

Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: A comprehensive review. *Journal of the American Geriatrics Society, 40,* 922–935. http://dx.doi.org/10.1111/j.1532-5415.1992.tb01992.x

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 4–70. http://dx.doi.org/10.1177/109442810031002

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE, 8*(7), e68967. http://dx.doi.org/10.1371/journal.pone.0068967

Webster, L., Groskreutz, D., Grinbergs-Saull, A., Howard, R., O'Brien, J. T., Mountain, G., . . . Livingston, G. (2017). Development of a core outcome set for disease modification trials in mild to moderate dementia: A systematic review, patient and public consultation and consensus recommendations. *Health Technology Assessment, 21,* 1–192. http://dx.doi.org/10.3310/hta21260

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods, 18,* 320–334. http://dx.doi.org/10.1037/a0032121

Wimo, A., & Prince, M. (2010). *World Alzheimer Report 2010: The global economic impact of dementia.* Retrieved from http://www.alz.co.uk/research/worldreport/

Yao, M., Zhu, S., Tian, Z. R., Song, Y. J., Yang, L., Wang, Y. J., & Cui, X. J. (2018). Cross-cultural adaptation of Roland-Morris Disability Questionnaire needs to assess the measurement properties: A systematic review. *Journal of Clinical Epidemiology, 99,* 113–122. http://dx.doi.org/10.1016/j.jclinepi.2018.03.011