Quality-of-life assessment in dementia: the use of DEMQOL and DEMQOL-Proxy total scores

Kia-Chong Chua[1] · Anna Brown[2] · Ryan Little[3] · David Matthews[3] ·
Liam Morton[1] · Vanessa Loftus[4] · Caroline Watchurst[1] · Rhian Tait[1] ·
Renee Romeo[1] · Sube Banerjee[5]

Purpose There is a need to determine whether health-related quality-of-life (HRQL) assessments in dementia capture what is important, to form a coherent basis for guiding research and clinical and policy decisions. This study investigated structural validity of HRQL assessments made using the DEMQOL system, with particular interest in studying domains that might be central to HRQL, and the external validity of these HRQL measurements.
Methods HRQL of people with dementia was evaluated by 868 self-reports (DEMQOL) and 909 proxy reports (DEMQOL-Proxy) at a community memory service. Exploratory and confirmatory factor analyses (EFA and CFA) were conducted using bifactor models to investigate domains that might be central to general HRQL. Reliability of the general and specific factors measured by the bifactor models was examined using omega (ω) and omega hierarchical (ω_h) coefficients. Multiple-indicators multiple-causes models were used to explore the external validity of these HRQL measurements in terms of their associations with other clinical assessments.
Results Bifactor models showed adequate goodness of fit, supporting HRQL in dementia as a general construct that underlies a diverse range of health indicators. At the same time, additional factors were necessary to explain residual covariation of items within specific health domains identified from the literature. Based on these models, DEMQOL and DEMQOL-Proxy overall total scores showed excellent reliability ($\omega_h > 0.8$). After accounting for common variance due to a general factor, subscale scores were less reliable ($\omega_h < 0.7$) for informing on individual differences in specific HRQL domains. Depression was more strongly associated with general HRQL based on DEMQOL than on DEMQOL-Proxy (-0.55 vs -0.22). Cognitive impairment had no reliable association with general HRQL based on DEMQOL or DEMQOL-Proxy.
Conclusions The tenability of a bifactor model of HRQL in dementia suggests that it is possible to retain theoretical focus on the assessment of a general phenomenon, while exploring variation in specific HRQL domains for insights on what may lie at the 'heart' of HRQL for people with dementia. These data suggest that DEMQOL and DEMQOL-Proxy total scores are likely to be accurate measures of individual differences in HRQL, but that subscale scores should not be used. No specific domain was solely responsible for general HRQL at dementia diagnosis. Better HRQL was moderately associated with less depressive symptoms, but this was less apparent based on informant reports. HRQL was not associated with severity of cognitive impairment.

✉ Kia-Chong Chua
k1072056@kcl.ac.uk

[1] Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF, UK

[2] School of Psychology, University of Kent, Canterbury, UK

[3] Croydon Memory Service, South London and Maudsley NHS Foundation Trust, London, UK

[4] Kingston Integrated Health and Social Care Community Mental Health Team for Older People, South West London and St George's NHS Trust, London, UK

[5] Brighton and Sussex Medical School, University of Sussex, Brighton, UK

Introduction

In dementia, as in other long-term conditions, 'adding life to years' is as important as 'adding years to life' [1]. The objective of assessing health-related quality-of-life (HRQL) is to be able to measure this. While medications used for people with dementia target cognitive and psychiatric symptoms, these symptoms do not give a complete picture of how illness can affect daily life and life quality [2]. HRQL measures are designed to include a broad range of domains in which impairments can occur and also where function and enjoyment can be maintained or even improved despite the progressive nature of dementia [3]. The broad view afforded by HRQL assessment is of particular value in multifaceted conditions with a broad range of physical, psychological and social impacts, such as dementia, to ensure that overall treatment benefits or harms are not missed [4].

How to obtain meaningful measurement of HRQL in dementia is an area of active research. In a recent systematic review that compared psychometric properties of HRQL measures for Alzheimer's disease and mixed dementia, the authors found 15 dementia-specific HRQL measures developed over the last 20 years [5]. The basis for measuring HRQL varies between instruments with different representations of what might be considered 'good' or 'bad' quality-of-life. There is a fundamental need to determine whether HRQL assessments in dementia capture what is important [6], to form a coherent basis for guiding research and clinical and policy decisions [7].

The first aim of this study was to explore structural validity of two HRQL measures, the DEMQOL and DEMQOL-Proxy, which relies on self- and informant report, respectively, for evaluating HRQL of people with dementia. These measures have shown good internal consistency, test–retest reliability and moderate evidence of validity in people with mild to moderate dementia for DEMQOL and mild, moderate and severe dementia for DEMQOL-Proxy [8]. Mulhern and colleagues [9] reported five domains in DEMQOL (cognition, negative emotion, positive emotion, social relationship and loneliness) and DEMQOL-Proxy (cognition, negative emotion, daily activities, positive emotion and appearance). In this study, we asked whether a coherent overall impression of a general phenomenon can emerge out of the complexities of multiple facets of HRQL. This entailed an investigation using unidimensional measurement models. Similar studies have also considered evidence of 'essential unidimensionality' [10] with a general factor in higher-order measurement models (e.g. second-order and bifactor models). These models retain substantive emphasis on a complex general phenomenon while recognising 'construct-relevant

multidimensionality' in which multiple domain-specific factors are necessary to reflect the inherent content diversity of complex constructs [11–13]. A key benefit from such a focus is empirical clarity in how well every DEMQOL and DEMQOL-Proxy item discriminates individual differences in overall HRQL, though this is more apparent in bifactor models than in second-order models because items in the latter load indirectly on the general factor [13, 14]. Multidimensional measurement models without a general factor (i.e. first-order correlated-factors model) would not aid this investigation. We thus considered only unidimensional, second-order and bifactor models.

The second aim was to investigate what might be important for quality-of-life around the time of a dementia diagnosis and how this might be captured by subscale and/or overall total HRQL scores. This question is primarily informed by investigations of content and face validity [15, 16]. The items in DEMQOL and DEMQOL-Proxy were generated in a process that included focus group interviews to assure rigorous coverage of relevant issues from the perspectives of people with dementia and their carers [17]. However, certain HRQL domains could matter more than others at different stages of the illness experience. Two measurement models in particular allow for an examination of the domains most central to the HRQL concept—the second-order and the bifactor model. In second-order models, how well domain factors load on the second-order general factor provides an indication of the relative importance of domain-specific functioning for general HRQL. In bifactor models, the amount of variation in item responses explained by the general vs specific factor provides a similar indication. If individual differences in a particular domain were fully explained by the general factor, this might suggest that the domain lies 'at heart' of the HRQL concept [18]. In this study, the domains in DEMQOL and DEMQOL-Proxy were examined for such insights as well as their implications on scoring practices.

The third aim of this study was to examine external validity by investigating the clinical relevance of individual differences in HRQL in terms of how they co-vary with clinically important outcomes in dementia. For this purpose, multiple-indicators multiple-causes (MIMIC) models with latent variables were used so that these conclusions were not affected by measurement unreliability.

Methods

Sample

The study participants were community-dwelling individuals and their carers referred to the Croydon Memory

Service, a service provided by the National Health Service (NHS) based in South London. This is a multidisciplinary and interagency team to generate early diagnosis in a timely manner, enabling choice and forward planning while people have capacity. It is designed to assess all incident cases in a given population. As well as diagnosis they provide information, and direct medical, psychological and social help to people with dementia and their family carers. They aim to prevent future crises by encouraging more effective and earlier help seeking and so reduce unwanted transition into care homes. The service model has been described in detail and has been subject to quantitative and qualitative evaluation [19, 20].

The subjects in this study were drawn from a series of consecutive cases who were referred to the service between December 2002 and June 2010. Cases were included in the analysis if, after a full multidisciplinary assessment (including physical examination, medical interview, laboratory and radiological investigations, neuropsychological assessment and mental state examination), they were given a formal clinical diagnosis of dementia using International Classification of Diseases (10th revision, ICD-10) diagnostic criteria [21]. They were excluded if they had not completed sufficient questions on the DEMQOL or DEMQOL-Proxy to allow the instruments to be scored. This sample therefore represents an analysis of routinely collected data of assessments of HRQL and other clinical assessments made at the time of first clinical diagnosis of dementia.

Measures

DEMQOL (28 items) and DEMQOL-Proxy (31 items) are interviewer-administered measures which obtain self- and informant reports of the HRQL of people with dementia [17]. Items inquire about 'feelings', 'memory' and 'everyday life' of the person with dementia in the last week. A four-point Likert scale (1 = a lot, 2 = quite a bit, 3 = a little, 4 = not at all) is used to collect responses. Reverse scoring is required for five items in DEMQOL/DEMQOL-Proxy so that higher overall total scores reflect better HRQL.

The data routinely collected by the memory service included measures of clinical symptoms in dementia as well as the HRQL. These included clinical assessments of cognition, depression, neuropsychiatric symptoms and dependence in activities in daily living. The Mini-Mental State Examination [MMSE, 22] is a screening tool for general cognitive impairment, with higher overall total scores (range 0–30) indicating better performance, and studies have reported evidence of structural validity [23], predictive validity and reliability [24–26]. The 15-item Geriatric Depression Scale [GDS-15, 27] is a screening

tool with higher overall total scores (range 0–15) indicating higher depression levels, and studies have reported evidence of concurrent validity [28–30] and diagnostic accuracy [30, 31]. The Neuropsychiatric Inventory [NPI, 32] is an assessment tool for frequency and severity of behavioural and psychological symptoms in dementia with higher overall scores (range 0–144) indicating poorer health, and studies have reported evidence of sensitivity to treatment-related changes [33, 34]. The Bristol Activities of Daily Living Scale [BADL, 35] is an assessment tool for functional decline among people with dementia in terms of their ability to carry out daily living activities independently with higher overall total scores (range 0–60) indicating more dependence, and studies have reported convergent validity and sensitivity to treatment-related changes [36, 37].

Analysis

Exploratory factor analysis

To establish a framework for the psychological constructs involved in HRQL measured by the DEMQOL and DEMQOL-Proxy, we conducted exploratory factor analysis (EFA) with bifactor orthogonal rotation [11]. One to six latent factors were considered in the EFA to explore domain themes of individual differences in HRQL response patterns of self-report (DEMQOL) and informant report (DEMQOL-Proxy), respectively. Eigenvalues and model fit were considered to aid factor retention decisions.

Confirmatory factor analysis

Among HRQL domains previously reported in the DEMQOL and DEMQOL-Proxy literature [9], some were absent from the bifactor EFA models in this study. While EFA results constitute 'absence of evidence' of the domain presence, 'evidence of absence' is needed to confirm that the domain does not emerge as a specific factor in bifactor confirmatory factor analysis (CFA). Signs of such 'factor collapse' [14] include (a) small and non-statistically significant factor loadings on a specific domain; (b) non-statistically significant factor variance of a specific domain. Model estimation may also fail to converge since 'factor collapse' implies over-extraction (i.e. hypothesising too many factors).

After the initial CFA, three types of model comparisons were made: (1) we first compared bifactor CFA models with and without the domain factors that were absent in EFA. Relative to more complex models (e.g. more domain factors), models that offered more parsimonious explanations of the sample data (e.g. fewer domain factors) would show poorer exact model fit. If the relative decline in model fit was trivial, this result would add 'evidence of

absence' to preceding investigations of factor collapse in the initial CFA; (2) having decided on a final bifactor CFA model for DEMQOL and DEMQOL-Proxy, respectively, we compared them with their nested second-order models. This alternative view of multidimensionality is a special case (i.e. nested model) of bifactor models [11, 38, 39], and thus second-order models can only fit the data worse. A recent simulation study has in fact demonstrated that the fit of bifactor model is unlikely to be challenged by second-order model and cautioned against relying on model comparison [40]; (3) we also included a comparison between bifactor CFA models and their strictly unidimensional counterparts to evaluate the extent in which the general HRQL factor was 'essentially unidimensional' [10] by comparing factor loadings on this general factor with those on the common factor of a strictly unidimensional model. These comparisons added to subsequent investigations aimed at informing whether individual differences in HRQL could be meaningfully interpreted with total scale scores and/or multiple subscale scores.

Reliability of model-based constructs

The CFA models imply ways in which DEMQOL (or DEMQOL-Proxy) scores could be used to reach conclusions about individual differences in HRQL. To see whether variation in overall total scores is mainly due to individual differences in general HRQL (i.e. good score reliability), we examined *factor saturation* using the omega hierarchical coefficient, ω_h [41, 42], which shows the percentage of variance in overall total scores that could be attributed to the target construct (general HRQL) in the presence of specific HRQL domains. As overall total scores have multiple sources of common variance (i.e. multidimensionality), reliability estimates would be more optimistic unless one of these sources of common variance is intended as the target construct using the ω_h [12]. We examined this issue using the omega (ω) coefficient, which shows the percentage of variance in overall total scores that could be attributed to all underlying factors (i.e. general *and* specific HRQL domains). Omega coefficients provide better estimates of measurement precision (reliability) than Cronbach's alpha [13], which conveys similar information, but is a special case of omega appropriate only for unidimensional factor models indicated by items with approximately equal factor loadings [43]. By modifying the calculation of omega coefficients [12, 18], we also investigated reliability of subscales in the context of bifactor multidimensionality.

Clinical associations with HRQL individual differences

To investigate the external validity of model-based HRQL constructs, we estimated their correlations with clinically relevant outcomes. We added to the CFA models four observed clinical covariates: cognitive functioning (MMSE), depression (GDS), neuropsychiatric symptoms (NPI), and dependence in daily life activities (BADL). We also explored potential differences due to gender, and whether HRQL assessments were fully or partially complete (e.g. self-report available for fewer than all 28 DEMQOL items). By working with the latent constructs emerging from DEMQOL/DEMQOL-Proxy, the associations were not affected by unreliability in HRQL assessments.

Modelling

All analyses were conducted in Mplus version 7 [44]. With a four-point Likert scale, DEMQOL and DEMQOL-Proxy responses were most appropriately treated as order categorical data [45]. The analyses were hence based on polychoric correlations rather than Pearson's correlations [46], and model parameters were estimated using the recommended diagonally weighted least squares (DWLS) estimator with robust standard errors, denoted 'weighted least squares means and variance adjusted' (WLSMV) in Mplus [47–49]. Overall model fit was evaluated in two ways. An *exact* fit between model predictions and observed data, within bounds of sampling error, would result in model Chi-square (χ^2) values that fail to reach statistical significance [50]. In addition to the Chi-square statistic, which is highly sensitive to sample size, a summary of *approximate* model fit was obtained. Approximate model fit is indicated by (1) low values of root mean square error of approximation [RMSEA, 51] where <0.10 is considered as acceptable and <0.05 as very good fit [52, 53]; and (2) high values of comparative fit index [CFI, 54] where >0.90 is considered as acceptable and >0.95 as very good fit [54, 55]. Modification indices, measured as improvement in exact model fit (or reduction in model χ^2 values) if constrained parameters are released, were used to inform modifications to the initial models. For models estimated with WLSMV, the DIFFTEST option in Mplus was required for model comparisons so as to obtain the correct Chi-square difference test ($\Delta\chi^2$) between models [44].

Results

Subjects

HRQL reports were obtained from 868 people with dementia and 909 informants. Details of the subjects with partially complete HRQL reports had slightly poorer health (e.g. GDS) than those for whom a full DEMQOL or DEMQOL-Proxy report was obtained (Table 1).

As the Croydon Memory Service was set up to facilitate early diagnosis for community-dwelling older adults, study participants were a sample of people who were in early stages of illness. While cognitive impairment based on MMSE scores is consistent with this (Table 1), NPI scores on average were below the means reported in clinical trials for mild to moderate dementia [e.g. 33]. BADL scores of the present sample also showed less functional decline than those reported in the BADL tool development study [35] which had people with more severe cognitive impairment.

EFA

With diverse outcomes in HRQL, a strictly unidimensional model was not tenable for DEMOQL ($\chi^2 = 4521.231$ ($df = 350$), RMSEA = .117 (90 % CI .114–.120), CFI = .686) and DEMQOL-Proxy ($\chi^2 = 6235.656$ ($df = 434$), RMSEA = .121 (90 % CI .119–.124), CFI = .681). Models with more domain factors gave better approximate fit even though model predictions did not reach an exact fit with DEMQOL and DEMQOL-Proxy data. Eigenvalues suggested a maximum of five factors might be considered for DEMQOL (10.540, 3.138, 1.690, 1.349, 1.187, 1.000) and a maximum of six factors for DEMQOL-Proxy (10.907, 3.277, 1.918, 1.581, 1.338, 1.207, 0.953). However, the ratio of the first two eigenvalues for DEMQOL (10.540 vs 3.138) and DEMQOL-Proxy (10.907 vs 3.277) suggested the presence of a strong general factor [13, 56].

For DEMQOL, we report the results of a bifactor EFA (**Model 1a**) that had a general HRQL factor and four domain-specific factors (supplementary Table 1). They were labelled as 'positive emotion' (POS: item 1, 3, 5, 6, 10), 'negative emotion' (NEG: item 4, 11, 12, 13), 'loneliness' (LON: item 8, 20) and 'worries about cognition' (COG: item 14, 15, 16, 17, 18, 19). Eleven DEMQOL items loaded saliently only on the general HRQL domain. For DEMQOL-Proxy, we report the results of a bifactor EFA (**Model 2a**) that had a general HRQL factor and five domain-specific factors (supplementary Table 2). They were labelled as 'positive emotion' (POS: item 1, 4, 6, 8, 11), 'negative emotion' (NEG: item 2, 3, 5, 7, 9, 10), 'worries about appearance' (APP: item 21, 22), 'worries about finance-related tasks' (FIN: item 23, 24, 25) and 'worries about social relationships' (SOC: item 27, 28, 29, 30). Eleven DEMQOL-Proxy items loaded saliently only on the general HRQL domain. Considerations that led to these final models included goodness of fit, interpretability of domain factor, fewer or weaker un-modelled cross-loadings and consistency with previous reports of multi-dimensionality [9, 17, 57].

Most of the HRQL domains reported in previous studies were replicated in the exploratory bifactor models of this study. However, the domain theme of 'worries about social functioning' (SOC) was absent from DEMQOL, whereas the domain theme of 'worries about cognition' (COG) was absent from DEMQOL-Proxy. These absent domains (SOC in DEMQOL and COG in DEMQOL-Proxy) formed the basis for investigating factor collapse in bifactor CFA models in the next stage.

CFA

Based on published findings [9], an additional domain 'worries about social relationships' (SOC: item 21, 22, 23, 24, 25, 26) was hypothesised, giving five specific domains (POS, NEG, LON, COG and SOC) alongside a general HRQL domain for DEMQOL (**Model 1b**). Similarly, an additional domain 'worries about cognition' (COG: item 12, 13, 14, 15, 16, 17, 18, 19, 20) was hypothesised, giving six specific domains (POS, NEG, APP, FIN, SOC, COG) alongside a general HRQL domain for DEMQOL-Proxy (**Model 2b**). With adequate approximate fit, bifactor CFA models for DEMQOL (RMSEA = .062 (90 % CI .059–.065), CFI = .918) and DEMQOL-Proxy (RMSEA = .058 (90 % CI .055–.061), CFI = .932) did not show evidence of factor collapse. The SOC domain in DEMQOL (supplementary Table 3) and COG domain in DEMQOL-Proxy (supplementary Table 4) had statistically significant factor variances and factor loadings.

In these CFA models, most items loaded saliently (≥ 0.3) on the general factor. The specific factor loadings of items tended to be weaker than their general factor loadings. In other words, general HRQL explained more variance in the item responses than specific domains did. Items that indicated 'positive emotion' (POS) in DEMQOL and DEMQOL-Proxy were an exception. Their factor loadings showed statistically significant but relatively weaker contributions towards general HRQL.

To further investigate the presence of specific HRQL domains, DEMQOL Model 1b was formally compared with a nested bifactor model without a SOC domain (**Model 1c**). Similarly, DEMQOL-Proxy Model 2b was compared with a bifactor model without a COG domain (**Model 2c**). DIFFTEST results show the decline in model fit was statistically significant for DEMQOL Model 1c relative to Model 1b ($\Delta\chi^2 = 172.023$, $df = 6$) and DEMQOL-Proxy Model 2c relative to Model 2b ($\Delta\chi^2 = 374.519$, $df = 9$). The subsequent stage of investigation proceeded with Model 1b for DEMQOL and Model 2b for DEMQOL-Proxy.

Next, DEMQOL Model 1b and DEMQOL-Proxy Model 2b were compared with their nested second-order models. While the second-order models had acceptable approximate model fit for DEMQOL (RMSEA = .065 (90 % CI .062–.068), CFI = .904) and DEMQOL-Proxy

Table 1 Demographic and clinical characteristics of the study group by completeness of HRQL rating by self-report (DEMQOL) and informant report (DEMQOL-Proxy)

	DEMQOL		DEMQOL-Proxy	
	Complete	Partial	Complete	Partial
Participants	756	112	679	230
Age[a]	78.7 (8.5)	77.9 (8.2)	78.8 (8.1)	79.3 (9.0)
	$n = 753$	$n = 112$	$n = 675$	$n = 230$
Gender				
Male	269	44	253	87
Female	487	68	426	143
Ethnicity				
White	657	86	580	191
Black	43	13	38	21
Asian	41	11	48	12
Unknown	15	2	13	6
ICD-10[b]				
AD	425	54	369	119
AD mixed	192	33	175	67
Vascular	84	15	84	30
Others	15	5	22	5
Unknown	40	5	29	9
MMSE[a]	21.1 (5.1)	20.4 (5.3)	20.8 (5.3)	19.5 (5.6)
(scores: 0–30)	$n = 756$	$n = 112$	$n = 679$	$n = 230$
GDS[a]	3.0 (2.6)	3.1 (2.6)	2.9 (2.7)	3.2 (2.6)
(scores: 0–15)	$n = 692$	$n = 105$	$n = 619$	$n = 198$
NPI[a]	12.7 (13.0)	12.3 (14.1)	12.1 (12.1)	15.2 (16.9)
(scores: 0–144)	$n = 684$	$n = 102$	$n = 668$	$n = 221$
BADL[a]	9.5 (9.2)	10.3 (9.4)	9.5 (9.2)	11.6 (9.9)
(scores: 0–60)	$n = 691$	$n = 105$	$n = 671$	$n = 225$

[a] Mean (SD). Rate of missing data varies across variables; valid sample size (n) is reported

[b] ICD-10 diagnosis: Alzheimer's disease, late/early onset (AD), Alzheimer's disease, mixed type (AD mixed), vascular dementia (vascular), others/unspecified (others), ICD code not known (unknown)

(RMSEA = .066 (90 % CI .064–.069), CFI = .905), they showed a statistically significant decline in exact model fit relative to their bifactor model counterparts (DEM QOL: $\Delta\chi^2 = 198.151$, $df = 18$; DEMQOL-Proxy: $\Delta\chi^2 = 369.875$, $df = 23$). Given that model fit comparisons have 'inherent statistical bias' in favour of bifactor models [40], this result was not surprising and highlighted that modelling and scoring approaches should be based on model utility.

In the final round of model comparisons, DEMQOL Model 1b and DEMQOL-Proxy Model 2b were evaluated against their strictly unidimensional counterparts (supplementary Table 3 and 4, respectively). The unidimensional models had poor model fit due to content diversity [58], but their factor loadings served as a reference for evaluating the impact on general factor loadings when items also load on additional domain factors as in the bifactor model. For these items, their factor loadings on the general factor were smaller than their factor loadings in the unidimensional model. This parameter distortion (due to un-modelled complexity in the latter) was expected, but only five had a magnitude of 0.10 or larger in the 28-item DEMQOL (e.g. item 10: 0.24 vs 0.45) and 31-item DEMQOL-Proxy (e.g. item 14: 0.57 vs 0.78), respectively. The extent of these differences between the general factor and the unidimensional common factor lends support to the view that general HRQL is essentially unidimensional.

Reliability

The general HRQL factor was a dominant influence on overall total scores in DEMQOL (supplementary Table 3: $\omega_h = 0.85$) and DEMQOL-Proxy (supplementary Table 4: $\omega_h = 0.88$). As there was more than one source of common variance underlying total scale scores (i.e. *GEN, POS, NEG, COG, LON, SOC* for DEMQOL; *GEN, POS NEG, APP, FIN, SOC, COG* for DEMQOL-Proxy), these would have led to more optimistic reliability estimates for

DEMQOL ($\omega = 0.96$) and DEMQOL-Proxy ($\omega = 0.96$). Going by ω estimates, all DEMQOL and DEMQOL-Proxy subscales showed excellent reliability ($\omega > 0.80$). When common variance in subscales was attributed to a general and specific source of influence, ω_h estimates showed that only 33–57 % of variation in subscale scores could be attributed to individual differences in specific HRQL domains. The POS domain was an exception. This subscale afforded excellent reliability in measuring individual differences in 'positive emotion' according to ω estimates and moderate reliability according to ω_h estimates in DEMQOL ($\omega = 0.86$ vs $\omega_h = 0.65$) and DEMQOL-Proxy ($\omega = 0.85$ vs $\omega_h = 0.69$).

External validity

Six covariates (MMSE, GDS, NPI, BADL, gender and complete/partial HRQL assessment) were added to the DEMQOL bifactor CFA Model 1b, generating Model 1d. DEMQOL-Proxy Model 2b was augmented with an identical set of covariates, generating Model 2d. The associations between HRQL and clinical outcomes (adjusted for gender differences and whether HRQL data were complete/partial) are presented in Table 2.

Higher levels of self-reported general HRQL (DEMQOL) were moderately associated with less depression (GDS). When rated by informants, general HRQL (DEMQOL-Proxy) had only weak associations with clinical outcomes. Males tended to have better general HRQL according to their informants.

Higher levels of 'positive emotion' (POS) according to self-report (DEMQOL) were moderately associated with less depression (GDS). In informant report (DEMQOL-Proxy), higher levels of POS were moderately associated with less dependence in daily living (BADL). In self-report (DEMQOL), less 'negative emotion' (i.e. higher levels of NEG) was associated with less depression. In informant report, less 'negative emotion' was associated with more neuropsychiatric problems (NPI).

In self-report, associations between 'worries about cognition' (COG) and clinical outcomes were weak. Less worries (i.e. higher levels of COG) were associated with more neuropsychiatric problems (NPI) and dependence (BADL). For DEMQOL-Proxy, a weak association was found between less worries and more dependence (BADL).

In self-report, a weak association was found between less 'worries about social relationship' (i.e. higher levels of SOC) and more depression (GDS). Males also fared worse in this domain. In informant report, less worries showed a weak association with less dependence (BADL).

'Loneliness' (LON), a domain unique to DEMQOL, showed little association with clinical outcomes, only that males showed less worries (i.e. higher levels of LON). Less

Table 2 External validity of HRQL measurements (standardised coefficients)

Model 1d	HRQL	POS	NEG	COG	SOC	LON
DEMQOL ($n = 724$)						
Gender	.04	.03	−.03	−.28	**−.41**	**.51**
Ax	−.11	−.02	−.01	−.02	−.41	−.26
MMSE	.00	.09	−.08	.05	−.02	.10
NPI	−.12	.00	.01	.19	.01	−.07
GDS	**−.55**	**−.49**	**−.32**	−.01	.28	−.09
BADL	.12	−.08	−.07	.18	−.08	.07

Model 2d	HRQL	POS	NEG	COG	SOC	APP	FIN
DEMQOL-Proxy ($n = 797$)							
Gender	**.39**	−.16	.05	−.09	−.02	.32	−.14
Ax	−.02	.32	.23	.00	.01	−.01	.00
MMSE	−.07	.01	−.03	−.08	−.08	−.13	.00
NPI	−.22	−.16	**−.48**	−.05	.07	.13	−.01
GDS	−.22	−.14	−.11	.07	.08	.13	.13
BADL	−.01	−.36	.05	.13	−.13	**−.42**	−.11

Model 1d: $\chi^2 = 1339.318$ ($df = 460$), RMSEA = .051 (90 % CI .048–.055), CFI = .903

Model 2d: $\chi^2 = 1599.858$ ($df = 550$), RMSEA = .049 (90 % CI .046–.052), CFI = .931

Gender with female as reference group; Ax: fully complete HRQL assessments served as reference group for comparing with partially complete

Standardised coefficients are italicised, if *unstandardised* coefficients were statistically significant. *Standardised* coefficients are bolded if they exceed a magnitude of 0.30

'worries about appearance' (i.e. higher levels of APP), a domain unique to DEMQOL-Proxy, were moderately associated with more dependence (BADL). 'Worries about finance-related tasks' (FIN), also unique to DEMQOL-Proxy, showed little association with clinical outcomes.

Discussion

HRQL as a multidimensional phenomenon in dementia

HRQL is commonly articulated as a complex phenomenon that needs to be understood in terms of multiple health-related domains. The complex nature of HRQL in dementia is apparent from previous factor analytic studies [9] which have shed light on multiple themes of individual differences in item response patterns of DEMQOL and DEMQOL-Proxy. Using bifactor model perspectives, this paper confirms earlier findings that items covering a diverse range of health-related domains can be combined to an

overall measure of HRQL in dementia. This finding aligns well with the substantive emphasis of HRQL assessments where the goal is to capture the overall balance of the impacts of diverse domains [59], particularly in treatment interventions that target broad outcomes [60]. By retaining strategic focus on general HRQL as the target construct, these analyses also show that some items (e.g. DEMQOL item 10) might be omitted from the assessment without affecting current levels of sensitivity in DEMQOL and DEMQOL-Proxy total scale scores to individual differences in a general complex phenomenon. This highlights the potential value of further analysis to consider the possibility of shorter versions of DEMQOL and DEMQOL-Proxy.

Furthermore, items from one domain, POS, had larger loadings on the domain factor than on the general factor, indicating that the POS-specific content was playing the more important role in responses to these items than the general HRQL factor. Reporting whether one had more 'positive emotions' or less 'worries' may also have different cognitive demands. Such influences have been reported in young children [61]. A recent population-based study has also reported an asymmetry of strong adverse reactions to deteriorations in health, alongside weak increases in well-being after health improvements [62]. Taken together, these issues may present challenges for overall HRQL scores to capture the variance of POS items, but this does not mean that positive emotion is not part of general HRQL. As POS items were the only items that required reverse-coding, the larger loadings on POS domain factor could also reflect this artefact [63–65]. Among studies investigating method effects [66–69], a multitrait–multimethod (MTMM) conceptual framework, comprising correlated-trait, correlated-uniqueness (CTCU) models, as well as correlated-trait, correlated-methods (CTCM) models, was employed to separate substantive content from method effects. While these analyses are beyond the scope of the present study, the orthogonality constraints of bifactor model framework provided the initial basis for speculating about the presence of potential method effects that are theoretically independent of individual differences in general HRQL [38]. However, these interpretations are post hoc, and thus preliminary, and a priori planned study designs that allow separating the substantive HRQL and common method effects (e.g. CTCU and CTCM models) are needed to reach a better understanding of this issue.

What matters in HRQL in dementia?

The bifactor EFA models in this study suggest that 'worries about social relationship' might be a core influence on how people with dementia rate their HRQL using the 28-item DEMQOL, whereas 'worries about cognition' might be central to how informants rate HRQL of people with dementia using the 31-item DEMQOL-Proxy. However, direct investigation of factor collapse using bifactor CFA models and model comparisons did not support the conclusion that 'worries about social relationship' were at the 'heart' of self-report HRQL in dementia. These latter analyses also did not support the conclusion that 'worries about cognition' were at the 'heart' of informant-rated HRQL. Such potential differences between self-report and proxy-report HRQL warrant continued investigation in light of the body of literature showing that self- and informant perspectives are influenced by different things [70–73]. With respect to social relationships, Lawton [74] suggested that social behaviour in people with dementia is 'a treatment goal that seems appropriate for an illness whose manifestations in general appear to represent estrangement from the external world'. As social functioning plays a pivotal role in the illness experience [75–78] as well as healthy ageing in general [79–82], factor collapse investigations using bifactor CFA such as those presented here may help shed light on whether social functioning could be considered a key clinical and policy focus when evaluating treatment interventions in dementia.

Subscales and overall total scores

It has been argued that subscale scores should be calculated because HRQL by definition is a multidimensional concept and respective domain scores might help clarify treatment impact [5, 83]. However, the current study suggests that after controlling for general HRQL, subscales in DEMQOL and DEMQOL-Proxy explain little more and have poor score reliability, and therefore should not be used.

This conclusion should not obstruct efforts to understand the specific ways in which treatment interventions have an impact on HRQL. Overall total scores can demonstrate whether treatment interventions may or may not be effective at a global level, amidst 'heterotypic continuity' [84] in which evidence of 'factor collapse' can show how different domains of the same underlying phenomenon may be central at different stages of illness [85].

Clinical associations with individual differences in HRQL

In line with prior research [4, 86], this study found that general HRQL had very little association with cognitive impairment and dependence in activities of daily living. Better HRQL was moderately associated with less depressive symptoms, but this was less apparent based on informant reports, possibly because depressive symptoms are less easily observed by informants [72].

It is worth noting that in development different items were found to work for self- and proxy report, so the 28 items in DEMQOL and 31 items in DEMQOL-Proxy are not identical. While this could have led to differences in construct validity of general HRQL, both measures do share four substantively similar domains (POS, NEG, COG, SOC). With content overlap in DEMQOL and DEMQOL-Proxy, there is also potential confusion over why some items that reflect negative emotion did not load on the NEG domain of DEMQOL, but they did load on NEG of DEMQOL-Proxy. In the context of bifactor models in which all domain factors are orthogonal, while negative emotions are integral elements of general HRQL, the elements of NEG domain carry 'incremental predic-tion' [87] which may reflect a form of negativity that is independent of self-report general HRQL. Following this logic, studies that have employed bifactor models have also shown that associations between these specific domains and external outcomes are not necessarily in the expected direction [e.g. 88, 89]. More definitive knowledge of the meaning of NEG (e.g. why it includes 'frustrated' and 'irritable' but not 'sad' and 'distress') and why NEG may differ in scope between self- and informant perspectives would require further research.

Study limitations

First, this study focussed on individuals with dementia around the time of diagnosis and is predominantly a sample of mild to moderately severe dementia. At more advanced stages of illness, HRQL may change for self-report and/or informant perspectives. The association between general HRQL and clinical outcomes may also vary by illness severity. The data reported here may not therefore be generalisable to populations with severe dementia or pos-sibly to those with more established dementia, in the years following diagnosis. Generalisability may be enhanced and selection bias minimised by the memory service being the setting for all diagnoses in a specific geographical area, as opposed to the subjects being drawn from a highly spe-cialised tertiary referral service.

Second, this was a convenience sample and potential bias from missing data cannot be ruled out. However, all cases where there were data on HRQL were included and all the data were collected as part of routine baseline clinical assessment, so it is not likely that selection bias is a particular problem. Also difficulties in obtaining a full HRQL report (DEMQOL/DEMQOL-Proxy) were only weakly related to illness severity.

Third, metric invariance [90], or the absence of non-uniform differential item functioning (DIF), had not been examined prior to testing the MIMIC models. While

MIMIC models aid the detection of uniform DIF, non-uniform DIF has to be investigated using multi-group factor analysis (MGFA). This presents two practical chal-lenges for the current study: (1) with six covariates, more than 12 models (at least 6 for DEMQOL and DEMQOL-Proxy, respectively) have to be estimated for MGFA; and (2) with covariates such as MMSE, NPI, GDS and BADL, widely accepted cut-off scores are needed before con-ducting MGFA. In this study, we leveraged on the flexi-bility of MIMIC models for a concurrent investigation with multiple covariates that vary in nature of measurement (categories/scores). Furthermore, simulation studies have demonstrated that MIMIC model approaches compare favourably with established methods (e.g. MGFA) for investigating uniform DIF [91–93]. In this study, we detected some DIF effects (supplementary Table 5), but they did not affect conclusions about external validity (supplementary Table 6). Taken together, these MIMIC models serve as a useful first-stage investigation for gen-erating hypotheses.

Finally, the themes that carry substantive relevance for HRQL in dementia may not be limited to the ones included in the DEMQOL measurement system. Given that other HRQL measures in dementia differ in content coverage, they may generate other findings about HRQL domains and what may matter at different stages of illness. DEMQOL is constrained by what is measurable on a Likert scale. Other measures and approaches may cover better other domains and determinants of what makes for quality-of-life in dementia, such as love or touch or time [94], which may be inaccessible to psychometrically based instruments.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

1. Clark, P. G. (1995). Quality of life, values, and teamwork in geriatric care: Do we communicate what we mean? *Gerontologist, 35*(3), 402–411.
2. Banerjee, S. (2007). Commentary on "Health economics and the value of therapy in Alzheimer's disease." Quality of life in dementia: Development and use of a disease-specific measure of health-related quality of life in dementia. *Alzheimers Dementia, 3*(3), 166–171.
3. Rabins, P. V., & Black, B. S. (2007). Measuring quality of life in dementia: Purposes, goals, challenges and progress. *International Psychogeriatrics, 19*(3), 401–407.
4. Banerjee, S., Smith, S. C., Lamping, D. L., Harwood, R. H., Foley, B., Smith, P., et al. (2006). Quality of life in dementia: More than just cognition. An analysis of associations with quality of life in dementia. *Journal of Neurology, Neurosurgery and Psychiatry, 77*(2), 146–148.
5. Perales, J., Cosco, T. D., Stephan, B. C., Haro, J. M., & Brayne, C. (2013). Health-related quality-of-life instruments for Alzheimer's disease and mixed dementia. *International Psychogeriatrics, 25*(5), 691–706.
6. Halvorsrud, L., & Kalfoss, M. (2007). The conceptualization and measurement of quality of life in older adults: A review of empirical studies published during 1994–2006. *European Journal of Ageing, 4*(4), 229–246.
7. Bakas, T., McLennon, S. M., Carpenter, J. S., Buelow, J. M., Otte, J. L., Hanna, K. M., et al. (2012). Systematic review of health-related quality of life models. *Health and Quality of Life Outcomes, 10*(1), 134.
8. Smith, S. C., Lamping, D. L., Banerjee, S., Harwood, R. H., Foley, B., Smith, P., et al. (2007). Development of a new measure of health-related quality of life for people with dementia: DEMQOL. *Psychological Medicine, 37*(5), 737–746.
9. Mulhern, B., Rowen, D., Brazier, J., Smith, S., Romeo, R., Tait, R., Watchurst, C., Chua, K. C., Loftus, V., Young, T., Lamping, D., Knapp, M., Howard, 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*(5), v–xv, 1–140.
10. Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*(3), 287–297.
11. Reise, S. P. (2012). Invited paper: The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696.
12. 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*(2), 129–140.
13. Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559.
14. 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*(2), 189–225.
15. Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., et al. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Medical Research Methodology, 10*, 22.
16. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research, 19*(4), 539–549.
17. Smith, S. C., Lamping, D. L., Banerjee, S., Harwood, R., Foley, B., Smith, P., Cook, J. C., Murray, J., Prince, M., Levin, E., Mann, A., & Knapp, M. (2005). Measurement of health-related quality of life for people with dementia: Development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technology Assessment, 9*(10), 1–93, iii–iv.
18. Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80*(4), 796–846.
19. Banerjee, S., & Wittenberg, R. (2009). Clinical and cost effectiveness of services for early diagnosis and intervention in dementia. *International Journal of Geriatric Psychiatry, 24*(7), 748–754.
20. Willis, R., Chan, J., Murray, J., Matthews, D., & Banerjee, S. (2009). People with dementia and their family carers' satisfaction with a memory service: A qualitative evaluation generating quality Indicators for dementia care. *Journal of Mental Health, 18*(1), 26–37.
21. 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 Journal of Geriatric Psychiatry, 22*(8), 782–788.
22. Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Minimental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198.
23. Rubright, J. D., Nandakumar, R., & Karlawish, J. (2015). Identifying an appropriate measurement modeling approach for the mini-mental state examination. *Psychological Assessment, 28*(2), 125–133.
24. Blake, H., McKinney, M., Treece, K., Lee, E., & Lincoln, N. B. (2002). An evaluation of screening measures for cognitive impairment after stroke. *Age and Ageing, 31*(6), 451–456.
25. Brayne, C. (1998). The mini-mental state examination, will we be using it in 2001? *International Journal of Geriatric Psychiatry, 13*(5), 285–290.
26. Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: A comprehensive review. *Journal of the American Geriatrics Society, 40*(9), 922–935.
27. Yesavage, J., & Sheikh, J. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist, 5*(1), 165–173.
28. Aikman, G. G., & Oehlert, M. E. (2001). Geriatric Depression Scale. *Clinical Gerontologist, 22*(3–4), 63–70.
29. Ferraro, F. R., & Chelminski, I. (1996). Preliminary normative data on the Geriatric Depression Scale-Short Form (GDS-SF) in a young adult sample. *Journal of Clinical Psychology, 52*(4), 443–447.
30. Herrmann, N., Mittmann, N., Silver, I. L., Shulman, K. I., Busto, U. A., Shear, N. H., & Naranjo, C. A. (1996). A validation study of the geriatric depression scale short form. *International Journal of Geriatric Psychiatry, 11*(5), 457–460.
31. Almeida, O. P., & Almeida, S. A. (1999). Short versions of the geriatric depression scale: A study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. *International Journal of Geriatric Psychiatry, 14*(10), 858–865.
32. Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J. (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology, 44*(12), 2308–2314.
33. Cummings, J. L., Ihl, R., Herrschaft, H., Hoerr, R., & Tribanek, M. (2013). Sensitivity to change of composite and frequency

scores of the Neuropsychiatric Inventory in mild to moderate dementia. *International Psychogeriatrics, 25*(3), 431–438.

34. Cummings, J. L., Tribanek, M., & Hoerr, R. (2014). Sensitivity to change of composite and frequency scores of the neuropsychiatric inventory in mild cognitive impairment. *International Psychogeriatrics, 26*(11), 1871–1874.

35. Bucks, R. S., Ashworth, D. L., Wilcock, G. K., & Siegfried, K. (1996). Assessment of activities of daily living in dementia: Development of the Bristol Activities of Daily Living Scale. *Age and Ageing, 25*(2), 113–120.

36. Sikkes, S. A., de Lange-de Klerk, E. S., Pijnenburg, Y. A., Scheltens, P., & Uitdehaag, B. M. (2009). A systematic review of Instrumental Activities of Daily Living Scales in dementia: Room for improvement. *Journal of Neurology, Neurosurgery and Psychiatry, 80*(1), 7–12.

37. Byrne, L. M., Wilson, P. M., Bucks, R. S., Hughes, A. O., & Wilcock, G. K. (2000). The sensitivity to change over time of the Bristol Activities of Daily Living Scale in Alzheimer's disease. *International Journal of Geriatric Psychiatry, 15*(7), 656–661.

38. Brown, A., & Croudace, T. (2015). Scoring and estimating score precision using multidimensional IRT. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment (a volume in the multivariate applications series)*. New York: Routledge/Taylor & Francis Group.

39. Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23*(1), 51–67.

40. Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41*(5), 407–422.

41. Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for omega(h). *Applied Psychological Measurement, 30*(2), 121–144.

42. Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and Mcdonald's ωH: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123–133.

43. Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104.

44. Muthén, L., & Muthén, B. (1998–2012). Mplus user's guide. In L. Muthén & B. Muthén (Eds.) (7th ed.). Los Angeles, CA. http://www.statmodel.com/ugexcerpts.shtml.

45. 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*(3), 354–373.

46. Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2008). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity, 44*(1), 153–166.

47. Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466–491.

48. Muthén, B., du Toit, S., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. http://www.statmodel.com/wlscv.shtml.

49. Savalei, V., & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology, 66*(2), 201–223.

50. Kline, R. B. (2010). Hypothesis testing. In T. Little (Ed.), *Principles and practice of structural equation modeling* (3rd ed., pp. 189–229). New York: Guilford Press.

51. Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173–180.

52. Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.

53. Reininghaus, U., McCabe, R., Burns, T., Croudace, T., & Priebe, S. (2011). Measuring patients' views: A bifactor model of distinct patient-reported outcomes in psychosis. *Psychological Medicine, 41*(2), 277–289.

54. Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246.

55. Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*(4), 557–585.

56. Reise, S. P., & Haviland, M. G. (2005). Item response theory and the measurement of clinical change. *Journal of Personality Assessment, 84*(3), 228–238.

57. Lucas-Carrasco, R., Lamping, D. L., Banerjee, S., Rejas, J., Smith, S. C., & Gomez-Benito, J. (2010). Validation of the Spanish version of the DEMQOL system. *International Psychogeriatrics, 22*(4), 589–597.

58. Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*(Suppl 1), 19–31.

59. 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*(11–12), 1249–1264.

60. Ebesutani, C., Reise, S. P., Chorpita, B. F., Ale, C., Regan, J., Young, J., Higa-McMillan, C., & Weisz, J. R. (2012). The Revised Child Anxiety and Depression Scale-Short Version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor. *Psychological Assessment, 24*(4), 833–845.

61. Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children—A cognitive developmental phenomenon. *Developmental Psychology, 22*(1), 37–49.

62. 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 and Medicine, 87*, 155–167.

63. Ebesutani, C., Drescher, C. F., Reise, S. P., Heiden, L., Hight, T. L., Damon, J. D., & Young, J. (2012). The loneliness questionnaire-short version: An evaluation of reverse-worded and non-reverse-worded items via item response theory. *Journal of Personality Assessment, 94*(4), 427–437.

64. Ebesutani, C., Drescher, C. F., Reise, S. P., Heiden, L., Hight, T. L., Damon, J. D., & Young, J. (2012). The importance of modeling method effects: Resolving the (uni)dimensionality of the loneliness questionnaire. *Journal of Personality Assessment, 94*(2), 186–195.

65. 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*(4), 394–404.

66. Tomás, 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: A Multidisciplinary Journal, 20*(2), 299–313.

67. Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thogersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment, 94*(2), 196–204.

68. Horan, P. M., DiStefano, C., & Motl, R. W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling, 10*(3), 435–455.

69. Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*(4), 810–819.

70. Black, B. S., Johnston, D., Morrison, A., Rabins, P. V., Lyketsos, C. G., & Samus, Q. M. (2012). Quality of life of community-residing persons with dementia based on self-rated and caregiver-rated measures. *Quality of Life Research, 21*(8), 1379–1389.

71. Moyle, W., Murfield, J. E., Griffiths, S. G., & Venturato, L. (2012). Assessing quality of life of older people with dementia: A comparison of quantitative self-report and proxy accounts. *Journal of Advanced Nursing, 68*(10), 2237–2246.

72. 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, 10*(5), 443–452.

73. Vogel, A., Mortensen, E. L., Hasselbalch, S. G., Andersen, B. B., & Waldemar, G. (2006). Patient versus informant reported quality of life in the earliest phases of Alzheimer's disease. *International Journal of Geriatric Psychiatry, 21*(12), 1132–1138.

74. Lawton, M. P. (1994). Quality of life in Alzheimer disease. *Alzheimer Disease and Associated Disorders, 8*(Suppl 3), 138–150.

75. Frick, U., Irving, H., & Rehm, J. (2012). Social relationships as a major determinant in the valuation of health states. *Quality of Life Research, 21*(2), 209–213.

76. 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*(4), 587–595.

77. 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*(2), 215–222.

78. 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*(4), 445–456.

79. Coyle, C. E., & Dugan, E. (2012). Social isolation, loneliness and health among older adults. *Journal of Aging Health, 24*(8), 1346–1363.

80. 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*(1), 3–16.

81. 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 and Medicine, 94*, 83–90.

82. 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*(4), 912–923.

83. Ettema, T. P., Droes, R. M., de Lange, J., Mellenbergh, G. J., & Ribbe, M. W. (2005). A review of quality of life instruments used in dementia. *Quality of Life Research, 14*(3), 675–686.

84. Holmbeck, G., Devine, K., & Bruno, E. (2010). Developmental issues and considerations in research and practice. In J. Weisz & A. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (2nd ed., pp. 28–39). New York, NY: Guilford Press.

85. Ebesutani, C., Smith, A., Bernstein, A., Chorpita, B. F., Higa-McMillan, C., & Nakamura, B. (2011). A bifactor model of negative affectivity: Fear and distress components among younger and older youth. *Psychological Assessment, 23*(3), 679–691.

86. Banerjee, S., Samsi, K., Petrie, C. D., Alvir, J., Treglia, M., Schwam, E. M., & del Valle, M. (2009). What do we know about quality of life in dementia? A review of the emerging evidence on the predictive and explanatory value of disease specific measures of health related quality of life in people with dementia. *International Journal of Geriatric Psychiatry, 24*(1), 15–24.

87. Ozer, D., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. In S. Fiske, A. Kazdin, & D. Schacter (Eds.), *Annual review of psychology* (Vol. 57, pp. 401–421). Palo Alto, CA: Annual Reviews.

88. Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80*(1), 219–251.

89. Yang, Y., Sun, Y., Zhang, Y., Jiang, Y., Tang, J., Zhu, X., & Miao, D. (2013). Bifactor item response theory model of acute stress response. *PLoS ONE, 8*(6), e65291.

90. Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3–4), 117–144.

91. Finch, H. (2005). The MIMIC model as a method for detecting DIF comparison with Mantel–Haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement, 29*(4), 278–295.

92. Willse, J. T., & Goodman, J. T. (2008). Comparison of multiple-indicators, multiple-causes- and item response theory-based analyses of subgroup differences. *Educational and Psychological Measurement, 68*(4), 587–602.

93. Woods, C. M. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research, 44*(1), 1–27.

94. Banerjee, S., Willis, R., Graham, N., & Gurland, B. J. (2010). The Stroud/ADI dementia quality framework: A cross-national population-level framework for assessing the quality of life impacts of services and policies for people with dementia and their family carers. *International Journal of Geriatric Psychiatry, 25*(3), 249–257.