

# A novel method of proxy reporting questionnaire based measures of health-related quality of life of people with dementia in residential care: a psychometric evaluation

Sarah C Smith  
AA Jolijn Hendriks  
Jemma Regan  
Nick Black

Department of Health Services  
Research and Policy, London School  
of Hygiene and Tropical Medicine,  
London, UK

**Background:** In research in residential care, health-related quality of life (HRQL) is usually measured using either observational methods or standardized questionnaires. DEMQOL-Proxy is a standardized questionnaire measuring HRQL of people with dementia and is usually reported by a family carer. However, not all residents have a family carer who visits often enough to act as a proxy.

**Objectives:** We evaluated the psychometric performance of DEMQOL-Proxy when reported on behalf of people with dementia in residential care by a “trained proxy” (DEMQOL-Proxy-TP).

**Participants:** We recruited a sample of 87 people with dementia living in care homes around the UK.

**Methods:** We used modern psychometric methods (based on the Rasch model) to evaluate DEMQOL-Proxy-TP (on behalf of 85 residents) in a cross-sectional study. We evaluated scale-to-sample targeting, ordering of item thresholds, item fit to the model and differential item functioning (sex, age, type of dementia), local independence, unidimensionality and reliability on the full set of items (31 items) and also a smaller item set (26 items).

**Results:** The smaller item set (DEMQOL-Proxy-TP-26) performed better than the original item set and was found to fit the model ( $p = 0.68$ ). Nevertheless, 17 items were found to have disordered thresholds, and 24 pairs of items showed local dependency (residual correlations  $>0.3$ ). There were also some areas where scale-to-sample targeting could be improved.

**Conclusion:** After resolving the identified anomalies, DEMQOL-Proxy-TP can provide adequate measurement of HRQL of people with dementia living in residential care, particularly when no family carer is available. This can be interpreted at the group level but is not yet robust enough for use at the individual level. Future work will compare these results with the psychometric performance of DEMQOL-Proxy reported by family carers and DEMQOL self-reported by the residents.

**Keywords:** dementia, DEMQOL-Proxy, health-related quality of life, Rasch Measurement Theory, trained proxy

## Introduction

Patient Reported Outcome Measures (PROMs) are standardized questionnaires widely used to measure the effect of interventions or treatment in both research and routine contexts.<sup>1</sup> In research in social care, such as residential care settings, using PROMs can be challenging, particularly in people with dementia. Outcomes such as health-related quality of life (HRQL) have sometimes been assessed using standardized PROMs such

Correspondence: Sarah C Smith  
Department of Health Services Research  
and Policy, London School of Hygiene and  
Tropical Medicine, 15-17 Tavistock Place,  
London WC1H 9SH, UK  
Tel +44 20 7927 2038  
Email sarah.smith@lshtm.ac.uk

as QOLAD<sup>2</sup> and DEMQOL.<sup>3</sup> Ideally, these questionnaires are self-completed by the residents, but often the nature of dementia makes self-completion impossible even when the questionnaire is interviewer administered. Nevertheless, it is important that a resident's opinion on his/her HRQL can be heard.

An alternative method is to use observational methods such as Dementia Care Mapping,<sup>4</sup> QUALIDEM<sup>5,6</sup> or QUALID<sup>7</sup> to obtain information about HRQL, and these can be advantageous in severe dementia or when language has deteriorated. However, these methods are labor intensive and potentially expensive to administer. Further, for some people in residential care, for example, residents who can communicate verbally and/or who have a family carer or other regular visitor who knows them well and can report a questionnaire on their behalf, these methods may not be necessary. Questionnaire measures have the advantage that they are not limited to observable behaviors and can ask respondents (or their proxy) directly about how they feel and the individual subjective impact of a particular health condition.

DEMQOL and DEMQOL-Proxy<sup>8–10</sup> are well-established PROMs to measure the HRQL of people with dementia. DEMQOL is self-reported by the person with dementia and is reliable and valid in people with mild/moderate dementia (Mini Mental State Examination score  $\geq 10$ ). DEMQOL-Proxy is proxy reported by a family carer or someone who knows the person with dementia well and sees him/her regularly. It is reliable and valid across the whole range of severity. The two instruments contain a common set of items, but also some unique items, and so are complementary rather than substitutable. In our previous work,<sup>11,12</sup> we have established robust new scoring algorithms for DEMQOL and DEMQOL-Proxy using Rasch-based psychometric methods, concluding that these measures are robust for use with people with dementia and their family carers.

DEMQOL/DEMQOL-Proxy have been previously used in residential care,<sup>3,13</sup> but no work has yet addressed the question of how well the instrument performs psychometrically when it is reported by proxies other than family carers. The question of who is the best alternative proxy after family carer has also not been widely considered. Not all people with dementia in residential care will have visitors, and those who do, may not visit frequently enough to be able to act as a proxy. Residential care staff may be optimistic in their report of HRQL as they are the provider of care. They may have a vested interest and may not therefore be an appropriate proxy either.

In our study, we investigated a new method of obtaining proxy reports of HRQL of people with dementia in residential care by using “trained proxies” (TPs) to report DEMQOL-Proxy on behalf of the residents (DEMQOL-Proxy-TP). TPs are professionals with experience in their existing professional roles of talking with people with dementia (dementia support workers, advocates and dementia advisors, possibly equivalent to level 5 or 6 of the European Qualification Framework, though we did not record this information directly). We conducted a psychometric evaluation of DEMQOL-Proxy-TP using modern psychometric methods, based on Rasch Measurement Theory.<sup>14,15</sup> They are more robust than traditional methods derived from Classical Test Theory.<sup>16,17</sup> Methods based on Rasch Measurement Theory enable scores to be derived that are truly interval (as opposed to ordinal) and independent of the sampling distribution of the items and of the sample (ie, they are said to be invariant). Where data fit the Rasch model, they create scores that have individual (rather than group level) standard errors, thus making the scores potentially usable for decision-making with individual respondents. Scores that are invariant and also truly interval have greater potential to measure meaningful change. We aim to investigate whether and to what extent reports from TPs can provide robust measures of HRQL and to identify potential anomalies with items and response scales so that these can be the focus of future work to improve DEMQOL-Proxy-TP.

## Methods

### Sample

A sample of 87 people with dementia living in residential care was recruited in a cross-sectional study. Participants were recruited from 16 care homes from three major chains: Anchor, Sunrise and Four Seasons. Care homes were selected to match the geographical availability of TPs. Homes were located in the North West, North East, East Anglia, East and West Midlands and South East. Residents were eligible for inclusion in the study if they had suspected dementia and had sufficient English to understand the consent process and talk to the TP. Written informed consent to take part in the study was obtained from each resident. For those residents who were not able to provide consent, written consent was obtained from a family/friend visitor or other representative (eg, care home staff member, health-care professional) by means of a Consultee Declaration form, provided the resident was willing to take part. The residents were 66% female, aged 66–99 years (mean age = 84.0, SD = 7.7), 89% White/White

British and mostly diagnosed with Alzheimer's dementia (39%) or vascular dementia (29%). A total of 31 TPs were able to complete a questionnaire for 85 of these participants. The study protocol was approved by the National Research Ethics Service Committee London (Reference: 14/LO/1835).

## Instrument

DEMQOL-Proxy<sup>8–10</sup> consists of 31 questions, each assessed on a 4-point Likert-type response scale: a lot, quite a bit, a little and not at all. The questions were derived from five conceptual domains: health and well-being (12 questions), cognitive functioning (9 questions), daily activities (6 questions), social relationships (2 questions) and self-concept (2 questions). Apart from emotion items, all items use the stem “how worried have you been about...”. There is also an additional question on overall quality of life, answered on a 4-point scale: very good, good, fair and poor. The items are scored according to a standard scoring algorithm<sup>18</sup> to produce an overall score where higher scores represent better HRQL.

## Procedure

TPs were recruited from the Alzheimer's Society and other independent advocacy agencies (Age UK, Advocacy Focus, Beth Johnson Foundation, Knowsley Pensioners Advocacy and Information Service). All TPs attended a training day prior to beginning the study. The training day covered how to explain the study, how to assess mental capacity and take consent and how to complete DEMQOL-Proxy on behalf of the resident. The training day also included specific open-ended time for TPs to consider and discuss the techniques they already used in their existing professional roles to engage with people with dementia. TPs were encouraged to share these ideas and suggestions with each other so that they all had clear ideas about possible ways to talk with the resident about the domains of HRQL covered by DEMQOL-Proxy. Each resident was assigned to a TP, who visited the resident on three occasions within a 2-week time frame. Most participating care homes had a single TP collecting data. In the care homes where two TPs shared the data collection, residents were allocated to a TP on a pragmatic basis, depending on the availability of the TP, the resident and their family visitor. The first two visits were to get to know the resident, introduce the study, assess the resident's mental capacity to consent and to obtain written informed consent. The third visit focused entirely on getting to know the resident using the TPs' existing professional skills and involved talking to the resident (sometimes using prompts such as pictures, objects in the room or the garden) about his/her life. At the

end of the third visit, the TP completed DEMQOL-Proxy-TP on behalf of the resident.

## Analysis

We conducted a Rasch analysis using RUMM2030 software to identify potential anomalies in the data indicating aspects of DEMQOL-Proxy-TP that were not working as intended.<sup>19</sup> Although all the items have the same 4-point Likert-type scale, the unrestricted (partial credit) model was used as this was an exploratory analysis and we wanted to be able to obtain an accurate impression of where anomalies lie and therefore to be able to address these in subsequent work. We examined the extent to which the scale was targeted to the sample, thresholds between each of the response options were ordered, each item fitted the model, different groups within the data (assuming the same amount of the construct being measured) did not show differences in scores (differential item functioning, DIF), items were independent of each other and items in the instrument represented a reliable, uni-dimensional construct.<sup>17,20</sup> All of the analyses were initially conducted for all 31 items and subsequently for a slightly smaller set of 26 items that excluded the positive emotion items as previous analyses<sup>11,12</sup> indicated that these items were conceptually different (personality trait items) and were unlikely to form part of a single dimension.

## Targeting

Scale-to-sample targeting concerns the match between the range of HRQL measured by the DEMQOL-Proxy-TP items and the range of HRQL in the sample of residents. This was evaluated by comparing the spread of person and item (threshold) locations.

## Ordering of item thresholds

We evaluated whether the response options were working as intended by a visual inspection of the item threshold map. As each item has four response options, there are three thresholds per item (point at which the probability of endorsing either of the two adjacent response options is the same) which should be ordered sequentially. Disordered thresholds can indicate where respondents have misunderstood or been unable to use response options consistently. Collapsing (or rescoring) the disordered thresholds can help to provide an indication of how response options can be improved.

## Item fit

There are three indicators of the fit of observed item data to the Rasch model: the fit residual (which needs to be within

the range of  $\pm 2.5$ ), the chi-square value and its (Bonferroni corrected) probability and the item characteristic curve (ICC) with the observed class interval mean scores (ie, the observed means of people grouped into a small number of class intervals of approximately equal numbers) plotted against their theoretical values based on the model. Fit residual and chi square are statistical indicators, whereas the ICC is a graphical indicator. No single piece of information can confirm the fit of an item to the model, and it is important therefore to consider all the evidence together.

### Differential item functioning

DIF is concerned with the extent to which the items work in the same way for different groups within the sample (eg, males and females), irrespective of their locations on the construct being measured. We used analysis of variance to evaluate whether observed (class interval) group means were the same, based on the premise that expected values are the same when the data fit the model. We defined groups as follows: sex of resident, age group (quartiles) and dementia type (Alzheimer's, vascular, other). There are two types of DIF: uniform DIF is indicated by a significant main effect for the group (eg, sex of resident), and nonuniform DIF is indicated by a significant interaction between the group and the class intervals. Uniform DIF can be corrected by treating problem items as different items for each group, known as "splitting" items (eg, separate items for males and females). Items showing nonuniform DIF may need to be investigated and/or removed from the item set.

### Local independence

The extent to which each item was independent of the others was evaluated by examining the residual correlation matrix (residual variance after the Rasch factor is accounted for). Pairs of items where the residuals were correlated  $>0.3$  were identified. Where items had residual correlations  $>0.3$ , we derived Rasch model-based scores with and without resolving for these items and correlated the two versions of the scores. If the correlation was high ( $>0.9$ ), we kept the initial scores from the initial set of items despite the high residual correlation in order to maintain content validity.

### Unidimensionality

The Rasch model assumes unidimensional data. This was evaluated using principal components analysis of the residuals. If there is no interpretable pattern in the residuals, then unidimensionality can be said to be supported.<sup>20</sup> Two subsets of four items were created from the highest and lowest loadings on the first principal component, and individual-level

paired *t*-tests were used to investigate whether the estimates for these two subsets differed significantly. We computed Wilson 95% CIs.<sup>21</sup>

### Reliability

Reliability was evaluated using the Person Separation Index (PSI), which is similar to Cronbach's alpha. A value  $>0.7$  is considered adequate.

### Rasch model-based (logit) scores and their benefit

Rasch model-based scores (logits) were generated for DEMQOL-Proxy-TP. The benefit of these scores was assessed by plotting them against the raw (original classically derived) scores. When the Rasch model-based scores are different to the raw scores, this will tend to give an ogive ("S"-shaped) curve.

## Results

For the full set of items (DEMQOL-Proxy-TP-31), overall chi square was significant ( $p < 0.001$ ), suggesting that the data did not fit the model. For the smaller item set (DEMQOL-Proxy-TP-26), overall chi square was nonsignificant ( $p = 0.68$ ) suggesting that the data fitted the model.

### Targeting

#### Original item set (DEMQOL-Proxy-TP-31)

The range of person locations was adequate to evaluate the items, but the range of item locations did not capture the full range of person locations (Figure 1); in particular, at the high end of the continuum, there were items lacking. There were also items at the low end of the continuum where no people were located. The positive mean value for persons (0.9 logits) indicates that the sample as a whole was located at a higher level of HRQL than the average of the scale.

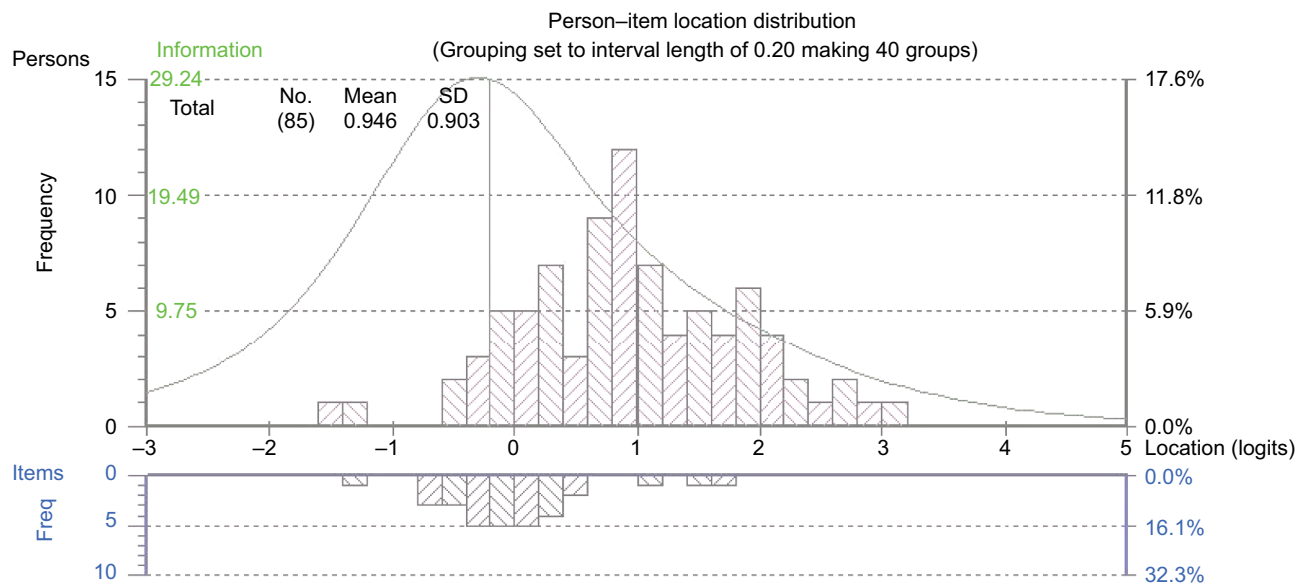
#### Smaller item set (DEMQOL-Proxy-TP-26)

For DEMQOL-Proxy-TP-26, person locations ranged from about  $-1.1$  logits to  $4.5$  logits ( $SD = 1.3$ ) and item locations ranged from about  $-1.1$  logits to about  $0.5$  logits ( $SD = 0.4$ ), which means that by leaving out the positive emotion items, for a relatively large part of the sample, items are missing that cover their level of HRQL (Figure 2).

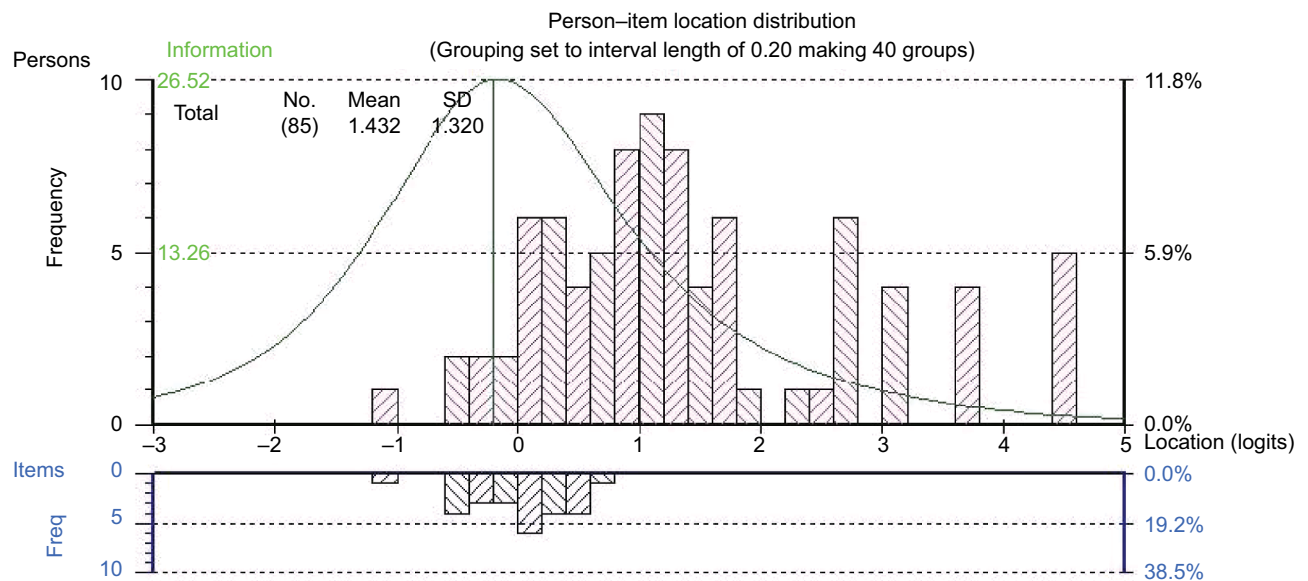
### Ordering of item thresholds

#### Original item set (DEMQOL-Proxy-TP-31)

Nineteen DEMQOL-Proxy-TP items showed response options not working as intended (disordered thresholds). This was particularly apparent for items asking about "worry



**Figure 1** Person-item location distribution for DEMQOL-Proxy-TP-31.



**Figure 2** Person-item location distribution for DEMQOL-Proxy-TP-26.

about cognitive function” and “worry about activities of daily living”, but also for other single items.

#### Smaller item set (DEMQOL-Proxy-TP-26)

For DEMQOL-Proxy-TP-26, we found 17 disordered thresholds. The pattern of items showing disordering was similar to that described above.

#### Item fit

##### Original item set (DEMQOL-Proxy-TP-31)

One DEMQOL-Proxy-TP item (felt full of energy) showed misfit to the model, based on its fit residual (2.982), chi-square value ( $\chi^2 = 31.21$ ,  $p < 0.01$  Bonferroni corrected) and fit to the ICC together (Table 1).



### Smaller item set (DEMQOL-Proxy-TP-26)

None of the 26 DEMQOL-Proxy-TP items in the smaller item set showed misfit to the model (Table 1).

## Differential item functioning

### Original item set (DEMQOL-Proxy-TP-31)

None of the items in the original item set showed DIF (for sex, age group or dementia type).

### Smaller item set (DEMQOL-Proxy-TP-26)

None of the items in the smaller item set showed DIF (for sex, age group or dementia type).

## Local independence

### Original item set (DEMQOL-Proxy-TP-31)

Thirty-six DEMQOL-Proxy-TP item pairs showed local dependency. Most values were in the range of 0.3–0.5, though a few were between 0.5 and 0.75 (Table 1). In particular, we found residual correlations among some of the “feelings” items (11) and also some of the “worry about cognitive function” items (15). Please note that Table 1 shows each residual correlation twice, once for each item in a pair (eg, given a residual correlation of 0.43 for frustrated/distressed, Table 1 shows 0.43 for frustrated and 0.43 for distressed).

### Smaller item set (DEMQOL-Proxy-TP-26)

In the smaller item set for DEMQOL-Proxy-TP, we found 24 residual correlations  $>0.3$ , mostly among the “worry about cognitive function” items (13) and to a lesser extent the “worry about activities of daily living” items (5) and the “feelings” items (5) (Table 1).

## Unidimensionality

### Original item set (DEMQOL-Proxy-TP-31)

The 31 items in the original item set of DEMQOL-Proxy-TP failed to form a unidimensional scale. The principal components analysis/*t*-test protocol showed that the two subsets of measurements differed significantly for 13.1% [7.5; 21.9] of the cases at the 5% level and 4.8% [1.4; 14.7] of the cases at the 1% level.

### Smaller item set (DEMQOL-Proxy-TP-26)

The items in DEMQOL-Proxy-TP-26 were not unidimensional. The two subsets of measurements differed significantly for 11.4% [6.1; 20.3] of the cases at the 5% level and for 1.3% [0.1; 9.9] of the cases at the 1% level. This suggests that the items do not represent a unidimensional scale.

## Reliability

### Original item set (DEMQOL-Proxy-TP-31)

For the original item set, we found  $\text{PSI} = 0.90$ , suggesting that the instrument discriminates well among people in terms of their HRQL (ie, high reliability).

### Smaller item set (DEMQOL-Proxy-TP-26)

For the smaller item set, we found  $\text{PSI} = 0.87$ , suggesting good discrimination among people in terms of their HRQL.

## Rasch model-based (logit) scores and their benefit

We derived Rasch model-based scores for the smaller item set (DEMQOL-Proxy-26-TP) because of its better performance. We rescored the 17 items with disordered thresholds (ie, collapsing response categories as necessary). In addition, we attempted to resolve for response dependencies, although it was not possible to successfully resolve all of these. The person location estimates of the original model and the best-resolved model correlated highly ( $r \geq 0.95$ ); therefore, we kept the initial estimates. Figure 3 shows the relationship between DEMQOL-Proxy-TP raw scores (simple sums across items) and Rasch model-based scores (logits) on the instrument. The S-shaped curve clearly indicates the benefit of the Rasch model-based scores: a 10-point increase in terms of raw scores corresponds to a variable amount of increase in terms of logits, dependent on the person's location on the raw score scale. Particularly at the extremes of the distribution, there is a notable difference.

## Discussion

Our findings suggest that TPs can report DEMQOL-Proxy-TP-26 on behalf of people with dementia in residential care, but that this estimate may lack precision. Although we identified a number of anomalies, overall the data fit the model and the resulting Rasch-based scores are an improvement on the original scoring of DEMQOL-Proxy-TP (demonstrated by the ogive curve in Figure 3). Therefore, where it is not possible to obtain a self-reported HRQL questionnaire and there is no appropriate family carer, TPs reporting DEMQOL-Proxy-TP may provide a means for people with dementia in residential care to have a voice about their HRQL. Scores can be interpreted at the group level but are not yet robust enough for use at the individual level.

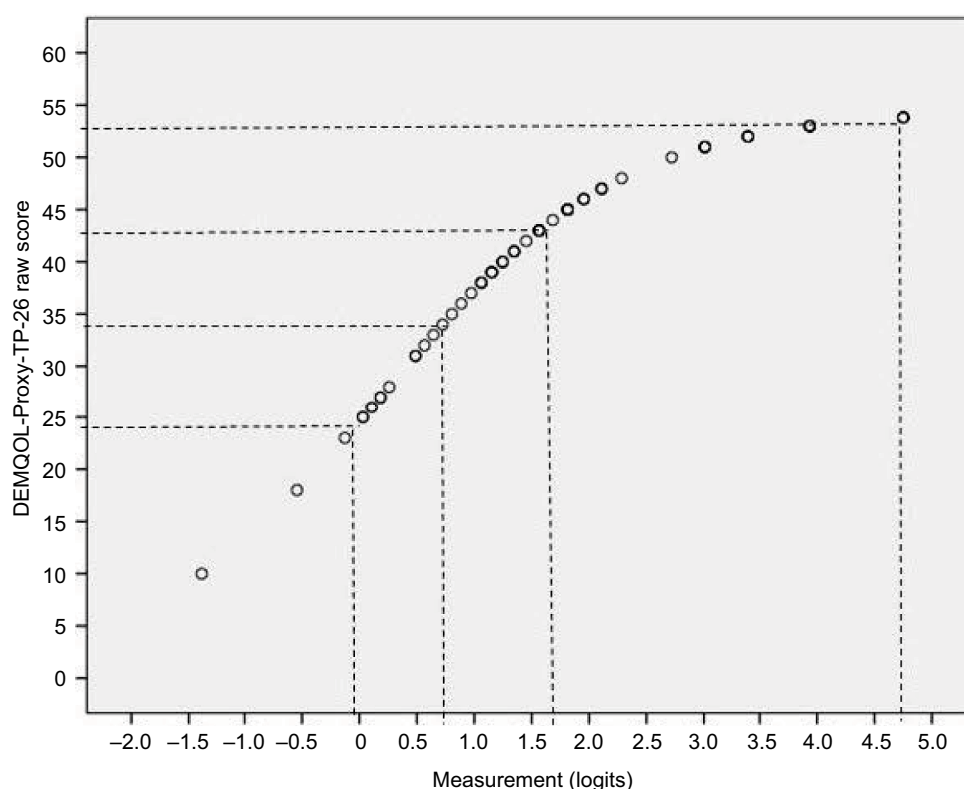
There are a number of items where the response options do not work as intended and also a large number of items showing local dependency. Together, these anomalies suggest that TPs may be using the instrument in a less nuanced way

Table 1 Diagnostic statistics

Item	DEMQOL-Proxy-TP-31						DEMQOL-Proxy-TP-26							
	Location	FitRes	ChiSq	p	ICC	DIF	ResCorr	Location	FitRes	ChiSq	p	ICC	DIF	ResCorr
Cheerful	0.535	0.133	0.32	0.85	++	No	0.43, 0.49, 0.46, 0.50							
Worried or anxious	0.223	-0.611	2.82	0.24	+	No	0.43	0.504	-0.240	0.59	0.75	++	No	0.39
Frustrated	0.174	-0.814	5.63	0.06	+/-	No	0.43, 0.46	0.424	-0.541	2.44	0.30	+	No	0.42, 0.41
Full of energy	1.644	<b>2.982</b>	<b>31.21</b>	<b>0.00</b>	—	No	0.43, 0.36, 0.68, 0.51							
Sad	0.320	-0.215	1.53	0.47	+	No	0.33	0.571	1.034	2.69	0.26	-	No	No
Content	0.409	0.413	0.18	0.92	++	No	0.49, 0.45							
Distressed	-0.234	-1.541	9.22	0.01	-	No	0.43, 0.33, 0.44, 0.39	0.019	-1.056	5.90	0.05	+/-	No	0.42, 0.44, 0.36
Lively	1.445	2.202	<b>21.10</b>	<b>0.00</b>	—	No	0.46, 0.68, 0.41							
Irritable	-0.222	-0.622	1.75	0.42	+	No	0.46, 0.44, 0.36	0.008	-0.041	1.20	0.55	+	No	0.41, 0.44, 0.34
Fed-up	0.358	-0.911	5.94	0.05	-	No	0.39, 0.36	0.615	0.383	2.27	0.32	+/-	No	0.36, 0.34
Things to look forward to	1.149	1.563	11.67	0.00	—	No	0.50, 0.51, 0.45, 0.41							
Memory in general	-0.350	-1.396	5.26	0.07	+/-	No	0.42, 0.35	-0.155	-0.956	0.78	0.68	++	No	0.38, 0.31
Forgetting happened long ago	-0.451	-0.831	3.30	0.19	+	No	0.42, 0.43, 0.41, 0.38, 0.39, 0.44	-0.260	-0.520	2.58	0.28	+	No	0.38, 0.43, 0.37, 0.33, 0.35, 0.42
Forgetting happened recently	0.166	-0.413	2.35	0.31	+	No	0.35, 0.43, 0.42, 0.46	0.379	-0.297	0.10	0.95	++	No	0.31, 0.43, 0.35, 0.38
Forgetting people's names	0.010	0.641	3.60	0.17	+/-	No	0.41, 0.42, 0.54	0.209	1.513	2.16	0.34	+/-	No	0.37, 0.35, 0.51
Forgetting where he/she is	-0.065	-0.616	1.70	0.43	+/-	No	0.34, 0.46	0.095	-0.506	1.26	0.53	+/-	No	0.41
Forgetting what day it is	0.127	-0.635	0.98	0.61	++	No	0.38, 0.46, 0.54, 0.31, 0.42	0.321	-0.220	0.97	0.62	++	No	0.33, 0.38, 0.51, 0.31
Thoughts being muddled	-0.034	-1.566	5.83	0.05	—	No	0.39, 0.34, 0.31, 0.61	0.156	-1.705	5.45	0.07	—	No	0.35, 0.55
Difficulty making decisions	-0.152	-1.400	2.51	0.29	-	No	0.44, 0.46, 0.42, 0.61	-0.002	-1.440	1.50	0.47	+	No	0.42, 0.41, 0.31, 0.55
Making himself/herself understood	-0.058	-0.593	1.53	0.46	+	No	No	0.158	0.141	0.26	0.88	+	No	No
Keeping himself/herself clean	-0.648	0.002	0.33	0.85	++	No	0.49	-0.513	0.143	0.23	0.89	++	No	0.50
Making himself/herself look nice	-0.493	-1.164	1.56	0.46	+	No	0.49	-0.421	-0.436	1.17	0.56	+/-	No	0.50
Getting from the shops	-1.339	-0.797	1.42	0.49	+	No	0.43	-1.130	-0.687	1.18	0.55	+	No	0.39
Using money to pay	-0.697	-0.328	0.95	0.62	+	No	0.76	-0.514	0.007	1.03	0.60	++	No	0.80
Looking after finances	-0.647	-0.142	0.93	0.63	+	No	0.76	-0.454	0.148	1.26	0.53	++	No	0.80
Things taking longer	-0.396	-0.266	0.22	0.90	++	No	0.37	-0.213	-0.070	3.17	0.20	-	No	0.32
Getting in touch with people	-0.146	-0.598	1.30	0.52	++	No	0.56	0.061	-0.116	0.81	0.67	++	No	0.51
Not having enough company	0.059	-0.553	1.46	0.48	+	No	0.56	0.258	0.203	0.88	0.64	+	No	0.51
Not able to help other people	-0.556	0.291	0.41	0.81	++	No	0.61	-0.381	-0.009	0.24	0.89	++	No	0.55
Not playing useful part	-0.379	-0.257	1.00	0.61	++	No	0.61	-0.196	-0.451	2.03	0.36	+	No	0.55
His/her physical health	0.247	0.008	0.69	0.71	+	No	0.37	0.460	0.423	4.47	0.11	-	No	0.32

**Notes:** Location: average item threshold location (logit); FitRes: fit residual ( $\geq 2.5$  in bold); ChiSq: chi-square value; p: chi-square probability; fit to the ICC: ++ = excellent, + = good, +/- = reasonable, - = marginal, — = bad; ResCorr: residual correlation. Chi-square tests that are statistically significant at family wise error rate  $\alpha = 0.01$  (Bonferroni corrected:  $p < 0.000323$ ) are in bold.

**Abbreviations:** ICC, item characteristic curve; DIF, differential item functioning.



**Figure 3** Relationship between raw scores and measurements (logits) for DEMQOL-Proxy-TP-26.

than intended. That is, TPs appear to be grouping response options together and using them as more gross-level categories. They also appear to be grouping some items together and responding to them more similarly than we would expect. In other contexts,<sup>12</sup> the response options appear to work reasonably well when used by people with dementia themselves and also by family carers. TPs appear to find it difficult to distinguish the response options (eg, a lot/quite a bit or quite a bit/a little) perhaps suggesting that their greater distance from the person with dementia means they cannot make such subtle distinctions between the meaning of these adjacent response options. Similarly, the high degree of local dependency suggests that TPs also find it difficult to distinguish between similar items. While they have gained enough understanding of the person with dementia to provide a general picture, they may not have enough knowledge to respond to subtly different questions and consequently these items tend to be answered too similarly. For the five items about worry about daily living and showing local dependency, this may also reflect the context of residential care and the lack of opportunity to do some of these things (eg, getting what he/she wants from the shops or using money to pay for things). Each of these questions asks whether the person is worried about doing these things, not whether they can

actually do them. However, it is possible that the TPs have interpreted these questions without the stem (ie, about how much the persons can do rather than how worried they are about whether they can do things). As a result, these items may have been grouped together and answered similarly on the basis of the residential care context and the likelihood that the resident does not have the opportunity to do some of these activities.

This is the first study to have developed an alternative method of proxy reporting of a PROM measuring HRQL of people with dementia in residential care. It is also the first to use psychometric methods to identify the impact of proxy reporting on the measurement properties of DEMQOL-Proxy in residential care. The features we have identified enable us to target further work more effectively. Future qualitative work should investigate in detail the TPs' perception and understanding of the items that show local dependency and problematic response options, particularly whether these are interpreted less subtly. The field notes completed by the TPs go some way toward this and suggest that the more general interpretation of some questions and response options may reflect the limited communicative ability of some of the residents. However, the field notes do not provide any insight about the specific response options that were harder



to use nor reasons why only certain questions were answered similarly. Our results also indicate that when used by TPs in residential care, there are improvements that can be made to the content of DEMQOL-Proxy-TP. In particular, more items that reflect the “better” end of the continuum need to be identified, and further testing in larger samples should evaluate whether items at the “worse” end of the continuum are indeed appropriate.

The main limitation of this study is the relatively small sample size. This will have reduced the precision of the estimates (larger standard errors) and will have made it more likely that outliers in the data will distort the estimates. However, the estimates reported here are broadly similar to those we have reported elsewhere using much larger samples<sup>12</sup> suggesting that the estimates in this study are reasonably robust. A sample size of 100 would give stable item calibrations  $\pm 0.5$  logit with 95% CI.<sup>22</sup> The results reported here are therefore robust enough for determining the problematic items and response options that need further investigation but require further replication in larger samples before the scores could be used in applied decision-making for either providers or individual residents. Further, our analysis is the first psychometric evaluation of DEMQOL-Proxy in the residential care context. We cannot therefore distinguish between findings that reflect the context of residential care and those that reflect the use of TPs. Future studies should tease this apart.

We have established that DEMQOL-Proxy-TP can provide reports of HRQL of people with dementia living in residential care, particularly when no family carer is available, but that these estimates are only to be used at the group level and are not yet robust enough for use with individual residents. Future work will need to compare these results with the psychometric performance of DEMQOL-Proxy reported by family carers and DEMQOL self-reported by the resident and to evaluate interrater reliability.

## Acknowledgments

The authors thank all the participants and TPs who took part in the study. This research was commissioned and funded by the Department of Health Policy Research Programme (Using Patient Reported Outcome Measures to Assess Quality of Life in Dementia, 0700071). The views expressed in this publication are those of the authors and not necessarily those of the Department of Health. The work reported in this paper was presented at the ISOQOL Conference 2016 as a poster. The poster's abstract was published in *Quality of Life*

*Research* (2016;25:1–196). The work was also presented at the Advances in Patient Reported Outcomes Research Conference 2016 as a poster.

## Disclosure

Dr Smith is one of the original developers of DEMQOL and DEMQOL-Proxy. The other authors report no other conflicts of interest in this work.

## References

1. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167.
2. Bosboom PR, Alfonso H, Almeida OP, Beer C. Use of potentially harmful medications and health-related quality of life among people with dementia living in residential aged care facilities. *Dement Geriatr Cogn Dis Extra*. 2012;2(1):361–371.
3. Maidment ID, Shaw RL, Killick K, et al. Improving the management of behaviour that challenges associated with dementia in care homes: protocol for pharmacy-health psychology intervention feasibility study. *BMJ Open*. 2016;6:e010279.
4. Kitwood T, Bredin K. A new approach to the evaluation of dementia care. *J Adv Health Nursing Care*. 1992;1(5):41–60.
5. Ettema TP, Dröes RM, de Lange J, Mellenbergh GJ, Ribbe MW. QUALIDEM: development and evaluation of a dementia specific quality of life instrument. Scalability, reliability and internal structure. *Int J Geriatr Psychiatry*. 2007;22(6):549–556.
6. Dichter MN, Schwab CG, Meyer G, Bartholomeyczik S, Halek M. Item distribution, internal consistency and inter-rater reliability of the German version of the QUALIDEM for people with mild to severe and very severe dementia. *BMC Geriatr*. 2016;16:126.
7. Weiner MF, Martin-Cook K, Svetlik DA, Saine K, Foster B, Fontaine CS. The quality of life in late-stage dementia (QUALID) scale. *J Am Med Dir Assoc*. 2000;1(3):114–116.
8. Smith SC, Lamping DL, Banerjee S, et al. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technol Assess*. 2005;9(10):1–93.
9. Smith SC, Murray J, Banerjee S, et al. What constitutes health-related quality of life in dementia? Development of a conceptual framework for people with dementia and their carers. *Int J Geriatr Psychiatry*. 2005;20:889–895.
10. Smith SC, Lamping DL, Banerjee S, et al. Development of a new measure of health-related quality of life for people with dementia: DEMQOL. *Psychol Med*. 2007;37:737–746.
11. Smith SC, Hendriks AAJ, Chrysanthaki T, Cano SJ, Black N. How can we interpret proxy reports of HRQL when it is no longer possible to obtain a self-report? Paper presented at: The ISOQOL Conference; October 21–24, 2015; Vancouver.
12. Hendriks AAJ, Smith SC, Chrysanthaki T, Cano SJ, Black N. DEMQOL and DEMQOL-Proxy: a Rasch analysis. *Health Qual Life Outcomes*. 2017;15(1):164.
13. Low LF, Brodaty H, Goodenough B, et al. The Sydney Multisite Intervention of LaughterBosses and ElderClowns (SMILE) study: cluster randomised trial of humour therapy in nursing homes. *BMJ Open*. 2013;3:e002072.
14. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: Danish Institute for Education Research; 1960.
15. Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978;43:561–573.
16. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol*. 2007;46:1–18.

17. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*. 2009;13(12):iii, ix–x, 1–177.
18. DEMQOL: dementia quality of life measure. Available from: <http://www.bsms.ac.uk/research/our-researchers/sube-banerjee/demqol/>. Accessed June 12, 2017.
19. Andrich D, Sheridan B. *RUMM 2030*. Perth: RUMM Laboratory Pty Ltd; 1997–2016.
20. Tennant A, Conaghan PG. The Rasch Measurement Model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum*. 2007;57:1358–1362.
21. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc*. 1927;22:209–212.
22. Linacre JM. Sample size and item calibration stability. *Rasch Meas Trans*. 1994;7.4:328.

## Patient Related Outcome Measures

### Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>

Dovepress

The journal is included in PubMed. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.