
Downloaded from: http://researchonline.lshtm.ac.uk/4648593/

DOI: https://doi.org/10.2147/PROM.S144000

Usage Guidelines:

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/
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. 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
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. They are more robust than traditional methods derived from Classical Test Theory. 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-Proxy8–10 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 algorithm18 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.19 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, unidimensional construct.17,20 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 analyses11,12 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 ±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. 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.

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

Smaller item set (DEMQL-Proxy-TP-26)
For DEMQL-Proxy-TP-26, we found 17 disordered thresholds. The pattern of items showing disordered was similar to that described above.

Item fit
Original item set (DEMQL-Proxy-TP-31)
One DEMQL-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 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 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

<table>
<thead>
<tr>
<th>Item</th>
<th>DEMQOL-Proxy-TP-31</th>
<th>DEMQOL-Proxy-TP-26</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Location</td>
<td>FitRes</td>
</tr>
<tr>
<td>Cheerful</td>
<td></td>
<td>0.535</td>
</tr>
<tr>
<td>Worried or anxious</td>
<td></td>
<td>0.223</td>
</tr>
<tr>
<td>Frustrated</td>
<td></td>
<td>0.174</td>
</tr>
<tr>
<td>Full of energy</td>
<td></td>
<td>1.644</td>
</tr>
<tr>
<td>Sad</td>
<td></td>
<td>0.320</td>
</tr>
<tr>
<td>Content</td>
<td></td>
<td>0.409</td>
</tr>
<tr>
<td>Distressed</td>
<td></td>
<td>−0.234</td>
</tr>
<tr>
<td>Lively</td>
<td></td>
<td>1.445</td>
</tr>
<tr>
<td>Irritable</td>
<td></td>
<td>−0.222</td>
</tr>
<tr>
<td>Fed-up</td>
<td></td>
<td>0.358</td>
</tr>
<tr>
<td>Things to look forward to</td>
<td></td>
<td>1.149</td>
</tr>
<tr>
<td>Memory in general</td>
<td></td>
<td>−0.350</td>
</tr>
<tr>
<td>Forgetting happened long ago</td>
<td></td>
<td>−0.451</td>
</tr>
<tr>
<td>Forgetting happened recently</td>
<td></td>
<td>0.166</td>
</tr>
<tr>
<td>Forgetting people’s names</td>
<td></td>
<td>0.010</td>
</tr>
<tr>
<td>Forgetting where he/she is</td>
<td></td>
<td>−0.065</td>
</tr>
<tr>
<td>Forgetting what day it is</td>
<td></td>
<td>0.127</td>
</tr>
<tr>
<td>Thoughts being muddled</td>
<td></td>
<td>−0.034</td>
</tr>
<tr>
<td>Difficulty making decisions</td>
<td></td>
<td>−0.152</td>
</tr>
<tr>
<td>Making himself/herself understood</td>
<td></td>
<td>−0.058</td>
</tr>
<tr>
<td>Keeping himself/herself clean</td>
<td></td>
<td>−0.648</td>
</tr>
<tr>
<td>Making himself/herself look nice</td>
<td></td>
<td>−0.493</td>
</tr>
<tr>
<td>Getting from the shops</td>
<td></td>
<td>−1.339</td>
</tr>
<tr>
<td>Using money to pay</td>
<td></td>
<td>−0.697</td>
</tr>
<tr>
<td>Looking after finances</td>
<td></td>
<td>−0.647</td>
</tr>
<tr>
<td>Things taking longer</td>
<td></td>
<td>−0.396</td>
</tr>
<tr>
<td>Getting in touch with people</td>
<td></td>
<td>−0.146</td>
</tr>
<tr>
<td>Not having enough company</td>
<td></td>
<td>0.059</td>
</tr>
<tr>
<td>Not able to help other people</td>
<td></td>
<td>−0.556</td>
</tr>
<tr>
<td>Not playing useful part</td>
<td></td>
<td>−0.379</td>
</tr>
<tr>
<td>His/her physical health</td>
<td></td>
<td>0.247</td>
</tr>
</tbody>
</table>

**Notes:** Location: average item threshold location (logit); FitRes: fit residual (>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; $p < 0.003223$) are in bold.

**Abbreviations:** ICC, item characteristic curve; DiF, differential item functioning.
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, 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\(^1\) suggesting that the estimates in this study are reasonably robust. A sample size of 100 would give stable item calibrations ±0.5 logit with 95% CI.\(^2\) 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**