Structural validity and internal consistency of the Qualidem in people with severe dementia

International Psychogeriatrics (2018), 30:1, 49–59 © International Psychogeriatric Association 2017
doi:10.1017/S1041610217001405

Structural validity and internal consistency of the Qualidem in people with severe dementia

Alexander M. M. Arons,[1] Roland B. Wetzels,[2,3] Sandra Zwijsen,[4] Hilde Verbeek,[5] Geertje van de Ven,[6] Teake P. Ettema,[7] Raymond T. C. M. Koopmans[2,8,9] and Debby L. Gerritsen[2,8]

[1]*Arons Consultancy, Nijmegen, the Netherlands*
[2]*Department of Primary and Community Care, Centre for Family Medicine, Geriatric Care and Public Health, Radboud University Medical Center, Nijmegen, the Netherlands*
[3]*Pleyade, Elderly Care Organization, Arnhem, the Netherlands*
[4]*Department of General Practice and Elderly Care Medicine, VU University Medical Center, Amsterdam, the Netherlands*
[5]*Department of Health Services Research CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, the Netherlands*
[6]*FWG, Research & Development, Utrecht, the Netherlands*
[7]*SHDH, Haarlem, the Netherlands*
[8]*Alzheimer Centre, Radboud University Medical Center, Nijmegen, the Netherlands*
[9]*Joachim en Anna, Center for Specialized Geriatric Care, Nijmegen, the Netherlands*

ABSTRACT

Background: Since its development, the Qualidem has had items that were considered unsuited for people with very severe dementia. This study attempted to investigate the applicability of all Qualidem items in people with all stages of dementia severity.

Methods: Four data sets that contained Qualidem observations on people with dementia were combined. Dementia severity was categorized based on the Global Deterioration Scale (GDS), with a dichotomization of very severe dementia (GDS 7) and others (GDS 1–6). Unidimensional latent-trait models (Mokken scaling) were estimated to fit the Qualidem responses in the overall sample and the dichotomized groups. Scalability was assessed using coefficients of homogeneity (Loevinger's H), while reliability was assessed with Cronbach's α and ρ.

Results: Combining the four databases resulted in 4,354 Qualidem measurements. The scalability of all scales was considered acceptable in the overall sample, as well is in the subgroups (all $H > 0.3$). Additionally, the reliability was good–excellent in the scales: "positive affect," "positive self-image," "care relationship," and "negative affect." Reliability was questionable–acceptable for "feeling at home," "social relations," "social isolation," and "restless tense behavior." Reliability was poor for "having something to do."

Conclusions: Statistical considerations allow using all Qualidem items in all dementia stages. Future research should determine balance of statistical- versus conceptual-based reasoning in this academic debate.

Key words: quality of life, dementia, validity, reliability, Mokken scaling

Introduction

Since the 1990s improving or stabilizing quality of life (QoL) has become a focus in dementia care. As such, QoL in dementia has been conceptualized and a number of instruments have been developed that aim to measure QoL (Lawton, 1994; Brod *et al.*, 1999; Logsdon *et al.*, 1999). Dementia is used as an umbrella term to refer to a number of different conditions that affect the brain, resulting in a decline in a number of cognitive functions, social functioning, and changes in mood and behavior. It is a progressive disorder that can be characterized by several stages ranging from (very) mild to (very) severe. In early to moderate stages of the disorder, patients' self-reports can be used to obtain information on QoL. However, as the degeneration continues/disease progresses, cognitive functions decline and information obtained from patients becomes increasingly difficult to interpret.

Correspondence should be addressed to: Roland B. Wetzels, MD, PhD Dept. of Primary and Community Care, Centre for Family Medicine, Geriatric Care and Public Health Radboud university medical center, P.O. Box 9101, 6500, HB, Nijmegen, the Netherlands. Phone: 024-3613338; Fax: 024-3619553. Email: roland.wetzels@radboudumc.nl. Received 3 Nov 2016; revision requested 19 Jan 2017; revised version received 22 Jun 2017; accepted 27 Jun 2017. First published online 4 September 2017.

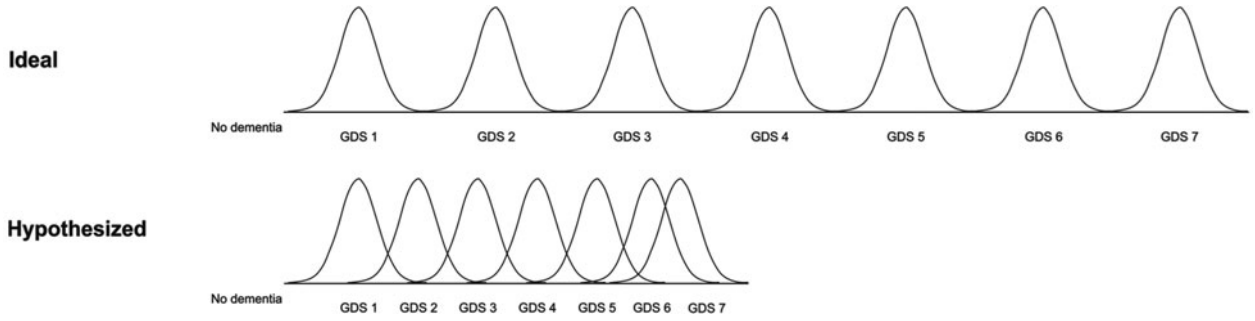

Figure 1. The amount of heterogeneity in GDS classifications.

An alternative to patients' self-reports on QOL in the more severe stages of dementia are assessments from proxies (informal or formal caregivers). The Qualidem (Ettema *et al.*, 2007a; Ettema *et al.*, 2007b) is a frequently used (Koopmans *et al.*, 2009; Verbeek *et al.*, 2010; Wetzels *et al.*, 2010a; 2010b; Zwijsen *et al.*, 2011; Oudman and Zwart, 2012; Schouten *et al.*, 2012; Wolf-Ostermann *et al.*, 2012; Teut *et al.*, 2013) dementia-specific proxy-rated QoL instrument developed specifically for people with dementia living in nursing homes/long-term care facilities. The instrument consists of 37 items that address observable behavior and has been regarded as valid and reliable in several studies (Ettema *et al.*, 2007a; 2007b; Bouman *et al.*, 2011). A review by Moyle and Murfield (2013) identified a number of instruments that can be used to measure QoL of people who have (very) severe dementia. They concluded that the Qualidem is applicable and practical. Additionally, a German study compared numerous QoL instruments and concluded that the Qualidem was one of their preferred instruments to assess QoL in people with dementia in shared housing arrangements (Gräske *et al.*, 2014). Nonetheless, Moyle and Murfield advocate a more thorough psychometric assessment of the instrument to be undertaken because of contradictory findings regarding the domain structure (Dichter *et al.*, 2011).

Currently, the instructions of the Qualidem instrument differ depending on the severity of dementia measured by the global deterioration scale (GDS) stage (Ettema *et al.*, 2007b), also called the Reisberg scale. This particular scale divides Alzheimer's disease into seven stages of ability, with stage 1 regarding no cognitive decline and stage 7 regarding very severe cognitive decline. In stages 2 through 6, all 37 Qualidem items are deemed applicable, while for stage 7 only 18 items are. The restrictions for severe dementia were imposed by the original authors because of the hypothesized inability to observe specific behaviors in people classified as GDS 7, which was confirmed after inspecting their initial data. For example, people classified as GDS 7 have lost all verbal abilities, and some basic psychomotor skills (Reisberg *et al.*, 1982), which causes difficulties in scoring items that rely on these abilities. This complicates the use of the Qualidem for monitoring persons with dementia over time in daily practice and in longitudinal studies; as people with dementia deteriorate during the follow up time, they might end up in GDS stage 7. With regard to analyzing data from longitudinal studies, two subgroups have to be created causing a loss of power for analyses. For these reasons some researchers choose to exclude participants who are in GDS stage 7 (Ven-Vakhteeva *et al.*, 2012). The fact that GDS classifications might be unstable over time (e.g. people who are at one time classified as GDS 7 can revert to GDS 6 (Wetzels *et al.*, 2010b)) complicates the use of the Qualidem further.

Additionally, the classification of the GDS might be too crude that causes people classified as GDS 7 to be a very heterogeneous group (Figure 1). Such heterogeneity might result in some scales to still be useful for comparisons and monitoring at the group level in the GDS 7 group, despite the fact that items in such scales cannot be used for all individuals. To overcome the above-mentioned difficulties, it would be ideal if all the Qualidem items could be used in the GDS 7 group. This would allow a more comprehensive assessment of QoL as well as improved applicability in daily practice and in longitudinal research.

The aim of the current study is to investigate the applicability of all Qualidem items in people with all stages of dementia severity.

Methods

Respondents

The current study combines data of participants from four different studies, all conducted in nursing homes throughout the Netherlands. The first study was the WAALBED-II-study, which investigated neuropsychiatric symptoms and physician's drug

https://doi.org/10.1017/S1041610217001405

prescription rates among people with dementia in nursing homes (Wetzels *et al.*, 2010b). The second study investigated the effects of a care program for managing challenging behavior of nursing home residents with dementia (Zwijsen *et al.*, 2011). The third study evaluated the effects of small-scale living facilities in dementia care on residents, family caregivers, and staff (Verbeek *et al.*, 2010). The fourth study evaluated the (cost-)effectiveness of dementia-care mapping, a cyclic person-centered intervention in nursing homes (van de Ven *et al.*, 2012; 2013). All studies had in common that they assessed QoL on an individual level over time using the Qualidem. It should be noted that each of the four studies was carried out in a nursing home setting, which makes the study samples comparable. Residents of Dutch nursing homes are mainly elderly people with an average age of 85 years. Those with dementia usually live in dementia special care units, which provide 24-hours a day nursing care, assistance with daily living, psychosocial, paramedical, and personal care. Dutch nursing homes employ multidisciplinary teams, which include nursing staff, an elderly care physician, a psychologist, a recreational therapist, a physical therapist, and a speech therapist. For the samples in the current study we report on age, sex, marital status, education, and severity of cognitive decline (GDS).

Instruments

The Qualidem is an often used and applicable (16, 17, 18) proxy-rated QoL instrument specifically developed for people with dementia in nursing homes, with 37 items intended to measure observable behavior related to QoL. It is filled in by nursing staff members. The items have the response options: "never," "seldom," "sometimes," and "often." Ettema et al. (2007b) identified nine homogenous subscales that they labeled: "care relationship," "positive affect," "negative affect," "restless tense behavior," "positive self-image," "social relations," "social isolation," "feeling at home," and "having something to do." Its validity and reliability have been deemed acceptable (Ettema *et al.*, 2007a; 2007b).

The GDS (Reisberg *et al.*, 1982) is a classification instrument indicating cumulative cognitive decline in dementia and can be used as an indicator to assess the severity of dementia. It has seven stages ranging from no memory complaints (stage 1) to loss of verbal skills, incontinence, and deterioration in simple psychomotor skills (stage 7).

Analyses

The four datasets (Verbeek *et al.*, 2010; Wetzels *et al.*, 2010b; Zwijsen *et al.*, 2011; van de Ven *et al.*, 2012) were combined and descriptive statistics were carried out. One dataset (Zwijsen *et al.*, 2011) was longitudinal and omitted some items at the first measurement, which made the calculation of four subscales impossible. For this reason, all data on that study's first measurement were excluded from the current study. All longitudinal datasets were restructured to provide several measurements of the same individual. A missing value analysis was performed on the Qualidem items using IBM SPSS version 20. Subsequently, missing data were imputed via a predictive mean matching algorithm (because Mokken scaling (see below) requires complete data records). All items with more than 10% missing data were excluded from further analyses, items with less than 5% missing data were considered to have data missing at random. Subsequently, the data file was imported into statistical package R (Team, 2005). To validate whether the previously identified data structure (Ettema *et al.*, 2007b) could be re-established in external datasets, the package "mokken" (Van der Ark, 2007) was used. This package can be used to test unidimensional latent trait models as identified by Mokken (1971). Mokken proposed two different models, the first is the monotone homogeneity model (MHM), the second is the double monotone homogeneity model (DMM). The application of the package "mokken" allows testing the goodness of fit for both these models.

The MHM assumption specifies whether or not a scale is cumulative. A cumulative scale implies the following: (1) the trait to be measured is a single trait and can be represented as a unidimensional continuum; (2) the probability of giving the positive response (saying "yes" to an item or response category of an item) does not decrease for subjects with increasing values of the latent trait; and (3) the probability of giving the positive response to each item and response category depends only on the value of the subject on the latent trait, and not on any other systematic influence. When these assumptions are met, a higher score on the scale implies more of the latent trait. The DMM is more restricted than the MHM. It has the added assumption that the order of the probabilities of the positive response to all items and response categories is the same for all subjects, regardless of their value on the latent trait. This assumption means that all subjects, independent of their scale values agree about the order of the "difficulty" of the items (the probability of saying "yes" to an item or response category). If a scale is found to conform to the DMM this means that in any population, any sample, any time period or any other experimental condition the ordering of subjects according to their scale value is the same (van Schuur, 2011).

https://doi.org/10.1017/S1041610217001405

In terms of model fit the scales are evaluated using Loevinger's coefficient of homogeneity H (Loevinger, 1947). H is defined as the ratio between observed error and expected error (under statistical independence). This coefficient can be defined between items ($\mathbf{H}_{ij}$), between an item and a scale ($\mathbf{H}_i$), and for a scale as a whole (**H**). These values can be interpreted as follows: 0–0.3 is insufficiently homogenous, 0.3–0.5 indicates weak homogeneity, and larger than 0.5 indicates strong homogeneity (Sijtsma and Molenaar, 2002). In addition, we investigate the "crit" values for the MHM and the DMM as defined by Sijtsma and Molenaar (2002). Crit values are calculated per item and take into account the number of comparisons with other items, the number of violations in these comparisons, the magnitude of these violations, the sum of the violations, the number of statistically significant violations, and a low $\mathbf{H_i}$. Sijtsma and Molenaar regard crit values below 40 as explainable from random fluctuations and values above 80 as serious violations of homogeneity. Scale reliability was assessed using Cronbach's α and ρ (Sijtsma and Molenaar, 2002). We investigated the nine scales that were originally identified by Ettema et al. (2007b) to see if we could replicate the results of the Mokken scaling in the entire group, in a group with GDS 1–6 and in a group with only GDS 7.

Finally, an item analysis was carried out to determine the percentage of answers per category per item. This was done to further illustrate potential differences between people with very severe dementia (GDS 7) and less severe dementia (GDS 1–6).

Results

Respondents

Combining the datasets of Verbeek *et al.* (2010), Wetzels *et al.* (2010b), Zwijsen *et al.* (2011), and van de Ven *et al.* (2012) resulted in a sample of 4,354 measurements, which consists of a combination of single and multiple measurements of 2,158 different respondents. The mean respondent age was 85.2 (SD 7.6, range 49–108) and 75.6% were female. Most individuals were classified as having moderate to (very) severe dementia (Table 1).

Missing value analysis

Item "Feels at home on the ward" had 2.9% missing values. All other items had less than 1% missing values.

Mokken scale analyses

There were two scales that adhere to the most restricted model, DMM. These were "positive self-image" and "social isolation." The scales "positive affect," "negative affect," "positive self-image," and "feeling at home," were considered strong scales ($\mathbf{H} > 0.5$) (Table 2), although "feeling at home" was not considered perfectly homogenous with a crit MHM value of 119 for the item "Feels at home on the ward." The reliability (ρ) of these scales ranged from 0.81 to 0.91, which was considered good to excellent. All the other scales were considered weakly homogenous ($0.3 \geq \mathbf{H} \leq 0.5$) and their reliability varied from 0.56 to 0.79, which was considered poor to acceptable. In the "Social relations" scale, there were two items that fell below the 0.3 threshold in all groups: "Takes care of other residents" and "Cuts oneself off from environment," both do not fit well with the other items in the scale. There was one item that weakened the "Care relationship" scale: This scale could be improved if the item "Appreciates help that he or she receives" would be removed.

There were minor deviations between the GDS 1–6 group and the GDS 7 group for all of the scales. However, none of the scales had insufficient homogeneity in either group. For both groups there was a single, albeit different, item that had a seriously high crit MHM value. These were "Feels at home on the ward" for the GDS 1–6 group, and "Appreciates the help he or she receives" for the GDS 7 group. The GDS 1–6 group had two additional items that were (suspicious of) violating homogeneity, while the GDS 7 group has four such items. In the GDS 7 group, the scale "Restless tense behavior" had weak homogeneity and it contains an item "Has tense body language" that falls below the 0.3 threshold.

Item analysis

The individual item analysis (Table 3) showed that especially for items 3 (Has contact with other residents), 18 (Takes care of other residents), 29 (Is on friendly terms with one or more residents), and 38 (Enjoys helping with chores on the ward) that the GDS 7 group had > 60% of the scores in the category "never." In the GDS 1–6 group, this was only applicable to item 38. These items fall in the scales of "Social relations" and "Having something to do."

Discussion

The aim of the current study was to investigate the applicability of all Qualidem items in people with all stages of dementia severity. Our main finding is that all the scales as identified by Ettema et al. (2007b) can be classified as having sufficient homogeneity using all items for the total

https://doi.org/10.1017/S1041610217001405

Table 1. GDS classifications of the combined samples

GDS classification		Total		Van de Ven *et al.*		Verbeek *et al.*		Zwijsen *et al.*		Wetzels *et al.*	
		N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
	1	6	0.14	6	1.09	0	0.00	0	0	0	0.00
	2	24	0.55	17	3.10	7	1.06	0	0.00	0	0.00
	3	80	1.83	37	6.74	34	5.14	9	0.47	0	0.00
	4	363	8.31	43	7.83	74	11.18	135	7.09	111	8.85
	5	1,205	27.58	103	18.76	216	32.63	565	29.67	321	25.60
	6	1,961	44.88	207	37.70	290	43.81	997	52.36	467	37.24
	7	694	15.88	135	24.59	28	4.23	178	9.35	353	28.15
Missing		21	0.50	1	0.20	1	0.20	20	1.10		
Total		4,354		549		650		1,904		1,252	
Gender											
Male		1,067	24.42	131	23.86	158	23.87	524	27.52	250	20.26
Female		3,301	75.56	417	75.96	504	76.13	1,380	72.48	1,002	79.74
Marital status											
Married		420	9.61	118	21.49	–	–	–	–	302	24.08
Unmarried cohabitation		6	0.14	6	1.09	–	–	–	–	0	0.00
Divorced/Unmarried		188	4.30	44	8.01	–	–	–	–	144	11.48
Widow		992	22.71	347	63.21	–	–	–	–	645	51.44
Other/Unknown		2,763	63.24		6.19					163	13.00
Education											
Low		834	19.09	242	44.08	–	–	–	–	592	47.21
Middle		99	2.27	7	1.28	–	–	–	–	92	7.34
High		27	0.62	0	0.00	–	–	–	–	27	2.15
Other/Missing		3,409	78.03	300	54.64	–	–	–	–	543	43.30
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age		85.14	7.57	84.14	6.44	82.94	7.23	83.61	7.20	89.07	7.29

group (GDS 1–7) and the GDS 1–6 group, but also for the subjects with very severe cognitive decline (GDS 7). Together with the fact that there were few missing values, which implies that all items could be answered meaningfully, this suggests that the statistical evidence points in a different direction than previous recommendations by Ettema et al.; although some scales could be improved by the omission of one or more items, the Qualidem's scalability is sufficient and, importantly, irrespective of the severity of dementia. In addition, the current findings do not indicate that the 18 items originally proposed by Ettema et al. (2007b) for measuring QoL in people with very severe dementia have higher scalability than the non-chosen items. This finding was corroborated by the individual item analysis, which showed that only a few items, especially on the scale "Social relations" show a marked difference between people with very severe dementia and less severe dementia.

These findings can be explained conceptually. Although some items may be difficult to rate in severe dementia, (e.g. "Wants to get off the ward" and "Finds things to do without help from others") the question is whether the difference between "no" and "not applicable" is relevant. In both cases, the resident's quality of life is expected to be lower than it would be if the response is "yes." This is not because severe dementia implies a lower quality of life, but because the resident has less resources and/or abilities that help him in achieving quality of life, which is in line with resource and capability-based approaches to quality of life (Finnema *et al.*, 2000; Ettema *et al.*, 2005). Nevertheless, for a scale to have content validity in severe dementia, it is essential that it also contains items that reflect behavior shown in very severe stages of dementia.

The scale "Care relationship" might be improved by omitting the item "Appreciates the help he or she receives." The scale "Restless tense behavior" is a weak scale overall and it contains one item "Has tense body language" that falls below the 0.3 threshold in the GDS 7 group. In the "Social relations" scale, there are two items that fall below the 0.3 threshold in all groups; "Takes care of other residents" and "Cuts oneself off from environment" both do not fit well with the other items in the

https://doi.org/10.1017/S1041610217001405

Table 2. Scale characteristics

		ALL GDS (N = 4,354, INCLUDES MISSINGS) SCALE H		RELIABILITY (CRONBACH'S α)		GDS 1–6 (N = 3,640) SCALE H		RELIABILITY (CRONBACH'S α)		GDS 7 (N = 693) SCALE H		RELIABILITY (CRONBACH'S α)	
ITEM NR. IN LIST	ITEM DESCRIPTION	ITEM HI	(ρ)	CRIT MHM	CRIT DMM	ITEM HI	(ρ)	CRIT MHM	CRIT DMM	ITEM HI	(ρ)	CRIT MHM	CRIT DMM
A. Care relationship		**0.43**	0.80			**0.45**	0.80			**0.43**	0.79		
4	Rejects help from nursing assistants	0.47	0.81	0	135	0.47	0.83	0	135	0.54	0.81	0	126
7	Is angry	0.51		0	100	0.51		0	96	0.53		11	100
14	Has conflicts with nursing assistants	0.55		0	142	0.56		0	132	0.56		0	152
17	Accuses others	0.39		0	180	0.41		0	170	0.39		0	126
24	Appreciates help that he or she receives	0.21		63	215	0.3		12	165	0.2		155	239
31	Accepts help	0.34		38	171	0.36		15	154	0.41		22	115
33	Criticizes the daily routine	0.41		0	209	0.44		0	160	0.4		0	137
B. Positive affect		**0.66**	0.91			**0.65**	0.90			**0.65**	0.90		
1	Is cheerful	0.68	0.91	0	0	0.66	0.91	0	0	0.73	0.91	0	41
5	Has a contented appearance	0.66		0	49	0.67		0	43	0.66		0	24
8	Is capable of enjoying things in daily life	0.65		0	0	0.64		0	0	0.64		3	0
10	Is in a good mood	0.73		0	0	0.72		0	0	0.72		0	32
21	Has a smile around the mouth	0.67		0	0	0.66		0	0	0.68		0	13
40	Mood can be influenced in a positive sense	0.56		0	54	0.54		0	50	0.61		58	37
C. Negative affect		**0.61**	0.79			**0.62**	0.80			**0.61**	0.77		
6	Makes an anxious impression	0.54	0.79	0	90	0.55	0.80	0	72	0.51	0.78	0	63
11	Is sad	0.65		0	61	0.65		0	45	0.67		0	46
23	Cries	0.65		0	66	0.65		0	48	0.65		0	45
D. Restless tense behavior		**0.36**	0.61			**0.36**	0.61			**0.38**	0.63		
2	Makes restless movements	0.42	0.61	0	71	0.41	0.62	0	64	0.48	0.65	20	127
19	Is restless	0.35		28	0	0.35		19	0	0.43		26	72
22	Has tense body language	0.3		18	77	0.32		0	68	0.24		59	165
E. Positive self-image		**0.67**	0.84			0.64	0.83			0.84	0.91		
27	Indicates that he or she would like more help	0.65	0.85	0	0	0.62	0.83	0	0	0.85	0.92	0	0

Table 2. Continued

		ALL GDS (N = 4,354, INCLUDES MISSINGS) SCALE H		RELIABILITY (CRONBACH'S α)		GDS 1–6 (N = 3,640) SCALE H		RELIABILITY (CRONBACH'S α)		GDS 7 (N = 693) SCALE H		RELIABILITY (CRONBACH'S α)	
ITEM NR. IN LIST	ITEM DESCRIPTION	ITEM HI	(ρ)	CRIT MHM	CRIT DMM	ITEM HI	(ρ)	CRIT MHM	CRIT DMM	ITEM HI	(ρ)	CRIT MHM	CRIT DMM
35	Indicates not being able to do anything	0.68		0	0	0.65		0	0	0.84		0	0
37	Indicates feeling worthless	0.68		0	0	0.66		0	0	0.81		0	0
F. Social relations		**0.34**	0.66			**0.3**	0.65			**0.34**	0.60		
3	Has contact with other residents	0.43	0.68	0	175	0.4	0.63	0	155	0.38	0.64	43	70
12	Responds positively when approached	0.4		0	86	0.34		0	98	0.43		38	34
18	Takes care of other residents	0.28		0	171	0.24		0	166	0.24		33	68
25	Cuts oneself off from environment	0.17		41	222	0.15		49	196	0.21		0	84
29	Is on friendly terms with one or more residents	0.41		0	194	0.37		0	178	0.41		10	60
34	Feels at ease in company of others	0.41		0	117	0.38		0	65	0.44		44	0
G. Social isolation		**0.44**	0.69			0.44	0.69			**0.41**	0.66		
16	Is rejected by other residents	0.46	0.69	0	0	0.46	0.69	0	0	0.46	0.67	0	0
20	Openly rejects contact with others	0.45		0	0	0.45		0	0	0.39		0	0
32	Calls out	0.4		0	0	0.39		0	0	0.37		0	0
H. Feeling at home		**0.5**	0.77			**0.51**	0.77			**0.46**	0.74		
13	Indicates that he or she is bored	0.53	0.81	0	139	0.52	0.81	0	104	0.58	0.76	0	144
28	Indicates feeling locked up	0.62		0	184	0.62		0	150	0.58		0	134
36	Feels at home on the ward	0.17		122	272	0.2		119	216	0.08		0	304
39	Wants to get off the ward	0.59		0	162	0.58		0	131	0.59		0	134
I. Having something to do		**0.42**	0.56			**0.39**	0.53			**0.5**	0.46		
26	Finds things to do without help from others	0.42	0.57	0	0	0.39	0.54	0	0	0.5	0.49	0	0
38	Enjoys helping with chores on the ward	0.42		64	0	0.39		53	0	0.5		0	0

Table 3. Individual item distributions of the QUALIDEM for people with GDS 1–6 and GDS 7

		INDIVIDUAL ITEM SCORES							
		GDS 1–6				GDS 7			
ITEM NR IN LIST	ITEM DESCRIPTION	0 (NEVER)	1 (RARELY)	2 (SOMETIMES)	3 (OFTEN)	0 (NEVER)	1 (RARELY)	2 (SOMETIMES)	3 (OFTEN)
A. Care relationship									
4	Rejects help from nursing assistants	19%	22%	17%	42%	17%	24%	14%	45%
7	Is angry	18%	34%	25%	23%	14%	30%	19%	37%
14	Has conflicts with nursing assistants	17%	24%	18%	40%	13%	16%	10%	61%
17	Accuses others	16%	20%	17%	48%	5%	5%	8%	82%
24	Appreciates help that he or she receives	3%	5%	20%	72%	15%	11%	27%	47%
31	Accepts help	1%	4%	17%	78%	4%	8%	22%	66%
33	Criticizes the daily routine	19%	23%	17%	41%	7%	7%	6%	80%
B. Positive affect									
1	Is cheerful	4%	13%	38%	45%	19%	19%	38%	24%
5	Has a contented appearance	5%	12%	34%	49%	13%	18%	34%	34%
8	Is capable of enjoying things in daily life	4%	11%	33%	52%	17%	23%	33%	28%
10	Is in a good mood	2%	8%	35%	55%	9%	15%	38%	38%
21	Has a smile around the mouth	5%	15%	37%	42%	15%	21%	37%	27%
40	Mood can be influenced in a positive sense	5%	10%	37%	48%	17%	15%	35%	33%
C. Negative affect									
6	Makes an anxious impression	17%	25%	19%	40%	17%	30%	15%	38%
11	Is sad	12%	30%	25%	34%	9%	24%	21%	46%
23	Cries	14%	16%	16%	55%	8%	11%	14%	67%
D. Restless tense behavior									
2	Makes restless movements	30%	17%	12%	41%	32%	19%	13%	37%
19	Is restless	31%	20%	18%	31%	27%	20%	15%	38%
22	Has tense body language	22%	24%	17%	36%	29%	27%	14%	30%
E. Positive self-image									
27	Indicates that he or she would like more help	18%	9%	8%	65%	6%	0%	3%	91%
35	Indicates not being able to do anything	18%	16%	14%	51%	8%	3%	3%	86%
37	Indicates feeling worthless	17%	11%	10%	61%	5%	3%	4%	88%

Table 3. Continued

		INDIVIDUAL ITEM SCORES							
		GDS 1–6				GDS 7			
ITEM NR IN LIST	ITEM DESCRIPTION	0 (NEVER)	1 (RARELY)	2 (SOMETIMES)	3 (OFTEN)	0 (NEVER)	1 (RARELY)	2 (SOMETIMES)	3 (OFTEN)
F. Social relations									
3	Has contact with other residents	17%	15%	25%	43%	64%	19%	10%	7%
12	Responds positively when approached	1%	4%	22%	73%	7%	12%	34%	48%
18	Takes care of other residents	51%	13%	15%	21%	84%	3%	2%	11%
25	Cuts oneself off from environment	25%	25%	16%	34%	39%	20%	8%	33%
29	Is on friendly terms with one or more residents	51%	11%	14%	24%	89%	3%	3%	4%
34	Feels at ease in company of others	3%	6%	26%	65%	12%	9%	26%	53%
G. Social isolation									
16	Is rejected by other residents	18%	21%	17%	45%	12%	17%	10%	60%
20	Openly rejects contact with others	17%	22%	17%	43%	11%	17%	10%	63%
32	Calls out	22%	16%	12%	50%	14%	14%	10%	61%
H. Feeling at home									
13	Indicates that he or she is bored	15%	15%	14%	55%	5%	3%	5%	87%
28	Indicates feeling locked up	42%	25%	20%	13%	5%	3%	2%	89%
36	Feels at home on the ward	4%	5%	19%	72%	6%	4%	19%	70%
39	Wants to get off the ward	20%	11%	10%	59%	6%	3%	3%	88%
I. Having something to do									
26	Finds things to do without help from others	53%	12%	14%	20%	86%	4%	3%	8%
38	Enjoys helping with chores on the ward	60%	12%	15%	13%	94%	3%	1%	2%

scale. There was one item that performed poorly in the "Care relationship" scale. This scale could be improved when the item "Appreciates help that he or she receives" would be removed. The scale "Feeling at home" might be improved by omitting the item "Feels at home on the ward."

The results of this study support from a statistical point of view the notion that all Qualidem items can be used to measure QoL in each GDS stage. This could have important implications for future studies as it will be easier, for example, to perform a power analysis, compare groups with different GDS scores, and use the Qualidem in longitudinal studies, thus greatly improving the instrument's applicability. For measurement in individuals, the applicability of all Qualidem items might still be open to debate, but the results of this study support the notion that its use at group-level measurement seems justified.

The results from the current study support the notion that in daily practice the Qualidem could be administered to all participants and all items should be administered. This is advantageous, as nurses will not have to worry or wonder which items are applicable or not. Moreover, when the Qualidem is translated and investigated in a different country (e.g. in Germany (Dichter *et al.*, 2011; Dichter *et al.*, 2013)), future studies will not have to perform analyses for separate groups, which makes the instrument much easier to apply and the results easier to interpret. We argue that at a group level all the Qualidem items can be applied in all levels of cognitive decline of people with dementia. This makes the Qualidem suitable for demonstrating effectiveness of interventions. Future studies could optimize the instrument further by performing in-depth analyses of which items deteriorate the validity and reliability of the instrument, and might be dropped. Additionally they could investigate the inclusion of items still under investigation.

A major strength of the current study is the number of respondents in the combined dataset, and if one agrees with the pooling of the datasets as has been performed in the current study, should thus supersede the results of previous studies such as Ettema *et al.* (2007a; 2007b) and Bouman *et al.* (2011). It is the largest dataset on which the Qualidem structure has been evaluated and thus provides the greatest confidence in its underlying domain structure, based on statistical conclusions. It could be considered a limitation that the study solely regards these statistical conclusions. Indeed, we are by no means ready to disregard the conceptual framework on which the Qualidem was built. Instead, we would like to invite fellow researchers and formal caregivers to join in this discussion of statistical versus conceptual-based reasoning.

Conclusion

In conclusion, the current study contributes to the literature regarding the validity and reliability of the Qualidem in particular, and to measurement in dementia in general. It suggests that measurement of quality of life in people with very severe dementia is possible.

Conflict of interest

None.

Description of authors' roles

All authors participated in the individual study concept and design; RBW, SZ, HV, GvdV, TPE collected the data; AMMA and RBW participated in acquisition of the different databases, AMMA merged the different databases and was responsible for the statistical design of the study and for carrying out the statistical analysis, AMMA wrote the draft article and RBW, SZ, HV, GvdV, TPE, RTCM, DLG assisted with writing the article.

Acknowledgments

This research was sponsored by VERENSO (the Dutch Association of Elderly Care Physicians and Social Geriatricians).

References

Bouman, A., Ettema, T., Wetzels, R., van Beek, A., de Lange, J. and Dröes, R. (2011). Evaluation of Qualidem: a dementia-specific quality of life instrument for persons with dementia in residential settings; scalability and reliability of subscales in four Dutch field surveys. *International Journal of Geriatric Psychiatry*, 26, 711–722.

Brod, M., Stewart, A. L., Sands, L. and Walton, P. (1999). Conceptualization and measurement of quality of life in dementia: the dementia quality of life instrument (DQoL). *The Gerontologist*, 39, 25–36.

Dichter, M. N. *et al.* (2013). Scalability and internal consistency of the German version of the dementia-specific quality of life instrument QUALIDEM in nursing homes-a secondary data analysis. *Health and Quality of Life Outcomes*, 11, 91. doi: 10.1186/1477-7525-11-91.

Dichter, M., Bartholomeyczik, S., Nordheim, J., Achterberg, W. and Halek, M. (2011). Validity, reliability, and feasibility of a quality of life questionnaire for people with dementia. *Zeitschrift Fur Gerontologie Und Geriatrie*, 44, 405–410.

Ettema, T. P., Dröes, R. M., de Lange, J., Mellenbergh, G. J. and Ribbe, M. W. (2007b). QUALIDEM: development and evaluation of a dementia specific quality

of life instrument. Scalability, reliability and internal structure. *International Journal of Geriatric Psychiatry*, 22, 549–556.

Ettema, T. P., Dröes, R.-M., de Lange, J., Ooms, M. E., Mellenbergh, G. J. and Ribbe, M. W. (2005). The concept of quality of life in dementia in the different stages of the disease. *International Psychogeriatrics*, 17, 353–370.

Ettema, T., Dröes, R.-M., de Lange, J., Mellenbergh, G. and Ribbe, M. (2007a). QUALIDEM: development and evaluation of a dementia specific quality of life instrument-validation. *International Journal of Geriatric Psychiatry*, 22, 424–430.

Finnema, E., Dröes, R.-M., Ribbe, M. and van Tilburg, W. (2000). A review of psychosocial models in psychogeriatrics: implications for care and research. *Alzheimer Disease & Associated Disorders*, 14, 68–80.

Gräske, J., Verbeek, H., Gellert, P., Fischer, T., Kuhlmey, A. and Wolf-Ostermann, K. (2014). How to measure quality of life in shared-housing arrangements? A comparison of dementia-specific instruments. *Quality of Life Research*, 23, 549–559.

Koopmans, R. T., van der Molen, M., Raats, M. and Ettema, T. P. (2009). Neuropsychiatric symptoms and quality of life in patients in the final phase of dementia. *International Journal of Geriatric Psychiatry*, 24, 25–32.

Lawton, M. P. (1994). Quality-of-life in Alzheimer-disease. *Alzheimer Disease & Associated Disorders*, 8, 138–150.

Loevinger, J. (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs: General and Applied*, 61, 1–49. DOI: 10.1037/h0093565.

Logsdon, R. G., Gibbons, L. E., McCurry, S. M. and Teri, L. (1999). Quality of life in Alzheimer's disease: patient and caregiver reports. *Journal of Mental Health and Aging*, 5, 21–32.

Mokken, R. J. (1971). *A Theory and Procedure of Scale Analysis*. The Hague, Netherlands: Mouton The Hague.

Moyle, W. and Murfield, J. E. (2013). Health-related quality of life in older people with severe dementia: challenges for measurement and management. *Expert Rev Pharmacoecon Outcomes Res*, 13, 109–122.

Oudman, E. and Zwart, E. (2012). Quality of life of patients with Korsakoff's syndrome and patients with dementia: a cross-sectional study. *Journal of the American Medical Directors Association*, 13, 778–781.

Reisberg, B., Ferris, S. H., de Leon, M. J. and Crook, T. (1982). The Global Deterioration Scale for assessment of primary degenerative dementia. *The American journal of psychiatry*, 139, 1136–1139.

Schouten, H. J., Knol, W., Egberts, T. C., Schobben, A. F., Jansen, P. A. and van Marum, R. J. (2012). Quality of life of elderly patients with antipsychotic-induced parkinsonism: a cross-sectional study. *Journal of the American Medical Directors Association*, 13, 82.e1–5.

Sijtsma, K. and Molenaar, I. W. (2002). *Introduction to Nonparametric Item Response Theory*. Thousand Oaks, CA: SAGE Publications, Incorporated.

Team, R. D. C. (2005). R: a language and environment for statistical computing. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2013. Available at: http://www.R-project.org.

Teut, M. *et al.* (2013). Effects and feasibility of an Integrative Medicine program for geriatric patients–a cluster-randomized pilot study. *Clinical interventions in aging*, 8, 953–961.

van de Ven, G. *et al.* (2012). Improving person-centred care in nursing homes through dementia-care mapping: design of a cluster-randomised controlled trial. *BMC geriatrics*, 12, 1–9.

van de Ven, G. *et al.* (2013). Effects of dementia-care mapping on residents and staff of care homes: a pragmatic cluster-randomised controlled trial. *PloS one*, 8, e67325.

Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, 20, 1–19.

van Schuur, W. H. (2011). *Ordinal Item Response Theory: Mokken Scale Analysis*. Thousand Oaks, CA: SAGE Publications, Incorporated.

Ven-Vakhteeva, J., Bor, H., Wetzels, R. B., Koopmans, R. T. and Zuidema, S. U. (2012). The impact of antipsychotics and neuropsychiatric symptoms on the quality of life of people with dementia living in nursing homes. *International Journal of Geriatric Psychiatry*, 28, 530–538.

Verbeek, H., Zwakhalen, S. M., van Rossum, E., Ambergen, T., Kempen, G. I. and Hamers, J. P. (2010). Dementia care redesigned: effects of small-scale living facilities on residents, their family caregivers, and staff. *Journal of the American Medical Directors Association*, 11, 662–670.

Wetzels, R. B., Zuidema, S. U., de Jonghe, J. F., Verhey, F. R. and Koopmans, R. T. (2010b). Course of neuropsychiatric symptoms in residents with dementia in nursing homes over 2-year period. *American Journal of Geriatric Psych*, 18, 1054–1065.

Wetzels, R., Zuidema, S., de Jonghe, J., Verhey, F. and Koopmans, R. (2010a). Determinants of quality of life in nursing home residents with dementia. *Dementia and Geriatric Cognitive Disorders*, 29, 189–197.

Wolf-Ostermann, K., Worch, A., Fischer, T., Wulff, I. and Gräske, J. (2012). Health outcomes and quality of life of residents of shared-housing arrangements compared to residents of special care units–results of the Berlin DeWeGE-study. *Journal of Clinical Nursing*, 21, 3047–3060.

Zwijsen, S. *et al.* (2011). Grip on challenging behaviour: a multidisciplinary care programme for managing behavioural problems in nursing home residents with dementia. Study protocol. *BMC Health Services Research*, 11, 41. Available at: https://doi.org/10.1186/1472-6963-11-41