

# Accepted Manuscript

Linguistic validation and reliability properties are weak investigated of most dementia-specific quality of life measurements - A systematic review

Martin Nikolaus Dichter, Christian G.G. Schwab, Gabriele Meyer, Sabine Bartholomeyczik, Margareta Halek



PII: S0895-4356(15)00377-7

DOI: [10.1016/j.jclinepi.2015.08.002](https://doi.org/10.1016/j.jclinepi.2015.08.002)

Reference: JCE 8954

To appear in: *Journal of Clinical Epidemiology*

Received Date: 20 February 2015

Revised Date: 20 July 2015

Accepted Date: 11 August 2015

Please cite this article as: Dichter MN, Schwab CGG, Meyer G, Bartholomeyczik S, Halek M, Linguistic validation and reliability properties are weak investigated of most dementia-specific quality of life measurements - A systematic review, *Journal of Clinical Epidemiology* (2015), doi: 10.1016/j.jclinepi.2015.08.002.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Title:** Linguistic validation and reliability properties are weak  
investigated of most dementia-specific quality of life measurements  
- A systematic review

**Martin Nikolaus Dichter<sup>1,2,\*</sup>, Christian G.G. Schwab<sup>1,2</sup>, Gabriele Meyer<sup>2,3</sup>, Sabine Bartholomeyczik<sup>2</sup>, Margareta Halek<sup>1,2</sup>**

<sup>1</sup> German Center for Neurodegenerative Diseases (DZNE), Stockumer Straße 12,  
58453 Witten, Germany

<sup>2</sup> School of Nursing Science, Witten/Herdecke University, Stockumer Straße 12,  
58453 Witten, Germany

<sup>3</sup> Institute for Health and Nursing Science, Medical Faculty, Martin Luther University  
Halle-Wittenberg, Germany

\*Corresponding author

E-Mail addresses:

MND: [Martin.Dichter@dzne.de](mailto:Martin.Dichter@dzne.de), phone: +49 2302926-253, fax: +49 2302926-239

CGGS: [Christian.Schwab@dzne.de](mailto:Christian.Schwab@dzne.de), phone: +49 2302926-258

GM: [Gabriele.Meyer@medizin.uni-halle.de](mailto:Gabriele.Meyer@medizin.uni-halle.de), phone: +49 345 557-4466

SB: [Sabine.Bartholomeyczik@uni-wh.de](mailto:Sabine.Bartholomeyczik@uni-wh.de), phone: +49 6221-7297159

MH: [Margareta.Halek@dzne.de](mailto:Margareta.Halek@dzne.de), phone +49 2302926-108

**Abstract****Objective**

For people with dementia, the concept of quality of life (QoL) reflects the disease's impact on the whole person. Thus, QoL is an increasingly used outcome measure in dementia research.

This systematic review was performed to identify available dementia-specific QoL measurements and to assess the quality of linguistic validations and reliability studies of these measurements (PROSPERO 2013: CRD42014008725).

**Study Design and Setting**

The Medline, CINAHL, EMBASE, PsychINFO and Cochrane Methodology Register databases were systematically searched without any date restrictions. Forward and backward citation tracking were performed based on selected articles.

**Results**

A total of 70 papers addressing 19 dementia-specific QoL measurements were identified; 9 measurements were adapted to non-origin countries. The quality of the linguistic validations varied from insufficient to good. Internal consistency was the most frequently tested reliability property. The majority of the reliability studies lacked internal validity.

**Conclusion**

QoL measurements for dementia are insufficiently linguistic validated and not well tested for reliability. None of the identified measurements can be recommended without further research. The application of international guidelines and quality criteria is strongly recommended for the performance of linguistic validations and reliability studies of dementia-specific QoL measurements.

**Keywords**

Quality of life, dementia, psychometric properties, reliability, linguistic validation, systematic review

**Running Title**

Dementia-specific quality of life measurements: a systematic review

**Word Count: 4988**

**Conflicts of interest**

The authors report no conflicts of interest or third-party financial support.

**What is new?****Key findings**

- Dementia-specific QoL measurements are often used in countries other than the origin; unfortunately, few measurements are linguistic validated, and their reliability properties are often unknown.
- Studies on psychometric properties often suffer from methodological shortcomings.

**What this adds to what is known?**

- This systematic review is the first to generate evidence on the lack of quality of linguistic validations and the reliability of dementia-specific QoL measurements.

**What is the implication, what should range now?**

- The application of international guidelines and quality criteria is strongly recommended for linguistic validations and reliability studies of dementia-specific QoL measurements.
- The application of these guidelines and quality criteria are recommended as prerequisites for publishing translated versions of QoL measurements and reliability studies.

## 1. Introduction

The estimated number of people with dementia worldwide was 44.4 million in 2013, and this number is estimated to increase to 135.5 million in 2050 [1]. Dementia as a syndrome results in progressive cognitive and functional declines that influence the affected individuals' performance of the activities of daily living and social behaviors. Quality of life (QoL) as a concept reflects the meaning that an individual attaches to the effects of the disease on him/her as a whole. Thus, it has become an important outcome in intervention studies, particularly psychosocial interventions, and an indicator of the quality of care of people with dementia [2, 3]. Dementia-specific QoL was first defined by Lawton as consisting of objective (e.g., behavioral competence and environment) and subjective (e.g., perceived QoL and psychological well-being) components [4]. Following Lawton's definition, other definitions have been suggested based on various theoretical concepts. All of the relevant theoretical concepts consider the subjectivity and multidimensionality of QoL [5, 6]. The lack of the concepts' theoretical clarity has resulted in several dementia-specific QoL measurements with heterogeneous operationalizations of the concept [7, 8]. Some measurements cover primarily functional and cognitive abilities, such as the Quality of life Alzheimer's disease scale [9], which is used to measure the health status rather than the QoL of people with dementia. By contrast, other measurements, such as the QUALIDEM [10], focus on the psycho-social domains of QoL. The self-rating of QoL by people with dementia is viewed as the gold standard method [3]. However, the reliability and validity of self-ratings are affected by deficits in memory, concentration, communication abilities, daily living capacity and impaired decision making, which occur progressively through the stages of the disease [11]. Therefore, proxy measures are recommended for longitudinal ratings and in advanced stages of the disease [2]. However, proxy-rating is also influenced by methodological difficulties, and the results are systematically lower than self-rated QoL [12]

and positively correlated with the raters' attitudes [13], burden [14, 15] and general life satisfaction [15]. In addition, the reliability and validity of proxy-ratings are affected throughout the entire course of dementia, because the observation of behaviors, moods, gesturing and facial expressions is challenging in the advanced stages of the disease. These theoretical and methodological difficulties emphasize the challenge of selecting the best Qol measurement for research and for dementia practice. These difficulties underline the relevance of a comprehensive and careful psychometric examination of dementia-specific Qol measurements [16].

Since the late nineties, nine systematic reviews of dementia-specific Qol measurements have been published [7, 8, 17-23] (Table 1). The number of included measurements increased with the publication year. The reviews report the perspective, content (subscales, items and response options), stage of dementia severity, and psychometric properties of the included measurements. Seven [7, 8, 17-19, 22, 23] of the nine reviews were based on a systematic literature search. Only one review included a detailed data extraction [8], and none reported the methodological quality of the included studies (Table 1).

Table 1: Reviews of dementia-specific Qol measurements

The majority of measurements that were included in previous systematic reviews were developed in native English-speaking countries [7, 8, 17-19, 22, 23]. Thus, the adaptation of the measurement in the context of increasing multinational research projects to gather comparable data on the Qol of people with dementia is crucial. Moreover, detailed analysis of the quality of linguistic validation processes in a review supports the selection of the most appropriate measurement in non-native English-speaking countries.

None of the former reviews systematically investigated different across national adapted versions of the Qol measurements.

Therefore, the objective of the present systematic review (PROSPERO 2013: CRD42014008725) is to assess the reliability of existing dementia-specific QoL measurements and to perform a critical appraisal of the quality of linguistic validations based on the recommendations of the AMSTAR tool [24].

## **2. Methods**

### *2.1 Search strategy*

In January and February 2014, a systematic search of the Medline [PubMed], CINAHL, EMBASE, PsychINFO [EMBASE] and Cochrane Methodology Register databases was performed without any date restrictions. In addition, potentially relevant publications known by the authors prior to the database search were considered. In a second step, forward and backward citation tracking of the included papers were performed using Web of Science and SCOPUS. The keywords used (and their combinations) are shown in Appendix B. The studies selected for inclusion were restricted to English or German language studies that primarily focused on the development, linguistic validation or reliability of dementia-specific QoL measurements. Studies and reviews that included people without any cognitive impairment were excluded. For the studies that reported further results, such as the predictors of QoL values or properties according to validity only, the study characteristics, methodological steps and results relevant to the objective of the current review were included.

### *2.2 Data extraction*

Each included study was categorized according to the study objective, sample inclusion and exclusion criteria, setting, proxy-rater inclusion and exclusion criteria, proxy training, sample size and properties of the dementia-specific QoL measurements (i.e., dementia severity,

perspective [self- or proxy-rating], time frame, subscales, number of items, response options and scoring procedure).

### *2.3 Quality assessment*

Two independent reviewers conducted the literature search and screening, data extraction and quality appraisal (MND, CGGS). In cases of disagreement, a third person assessed the paper, and the disagreements were discussed until a consensus was reached. To assess the quality of linguistic validation (called cross-cultural adaptation), each stage of the process recommended by Beaton et al. [25] was assessed. This process covers initial translation, synthesis, back translation, expert committee review and pilot testing of the preliminary translation (Appendix C). According to Acquadro et al. [26] we use the more precisely term of linguistic validation because cross cultural adaptation includes the translation (linguistic validation) and evaluation of the psychometric properties of a measurement. As recommended by Beaton et al. [25], and in accordance with Uysal-Bozkir et al.'s [27] procedure, the overall quality of the linguistic validation was rated as follows: 'good' if 4 to 5 of the steps of linguistic validation were rated as positive, 'moderate' if 2 to 3 of the steps were positively assessed and 'insufficient' if 0 to 1 of the steps were rated as positive. In cases in which no information of the linguistic validation was reported or the information was not available, no overall conclusion was reached. The quality of the internal consistency, inter-rater reliability and test-retest-reliability was assessed according to the criteria of the appraisal tool for studies of diagnostic reliability (QUAREL) by Lucas et al. [28] and the criteria for psychometric properties of Health Status Questionnaires by Terwee et al. [29] (Appendix D).

## **3. Results**

### *3.1 Search results*

Following the removal of 1943 duplicate references, the comprehensive database search resulted in 3659 references. Application of the inclusion and exclusion criteria led to the exclusion of an additional 3552 references, and subsequent screening of the titles and abstracts revealed 107 eligible references for full-text screening. A total of 59 papers remained for the review. Forward and backward citation tracking based on the reference lists and citations of the included 59 papers resulted in an additional 1319 publications after deleting 1423 duplicates. Screening revealed 11 additional papers for inclusion, resulting in a total of 70 included papers (Figure 1) addressing 19 dementia-specific QoL measurements (Appendix A).

Figure 1: A flow diagram of the literature search and study selection

### *3.2 Measurements assessing quality of life*

Of the 19 included dementia-specific QoL measurements, three were based on self-rating scores (BASQID [30], DQoL [11] and MCQ [31]); seven were based on self- and/or proxy-rating (AAIQOL [32], CBS QoL [33], DEMQOL [21], H.I.L.DE. [34], QoL-AD [9], QoL-AD Nursing home (NH) version [35], QOLAS [36]); and nine were based on proxy-rating only (ADRQL [37], CDQLP [38], DCM [39], OQOLD and OQOLDA [40], PWP-CIP [41], QoL-D [42], QUALID [43], QUALIDEM [10] and Vienna List [44]). The characteristics of the identified measurements (Appendix A), their reliability properties and information about their linguistic validations are summarized in Table 2. Detailed results of the quality appraisal for each of the reliability studies and the linguistic validations are shown in Appendices E and F.

Table 2: Results of the linguistic validation and reliability evaluation of dementia-specific QoL measurements

### *3.3 Self-rated Qol measurements*

#### BASQID

The BASQID is a self-rating measurement for people with mild to moderate dementia, and it is only available in its original British version [30, 45].

The internal consistency was well investigated in one study, with sufficient results [45]. One study with methodological weaknesses evaluated the test-retest reliability and demonstrated satisfactory results [30].

#### DQol

The DQol is a self-rating measurement for people with mild to moderate dementia. It was developed in the US [11] and has been used in Great Britain [46], Taiwan [47], France [48], Spain [49] and Germany [50]. For two versions, no linguistic validation was reported [21, 48, 51]; for the other versions, the quality of these validation processes was between insufficient [47, 49] and good [50]. The internal consistency of the original US version was investigated six times in a US sample [11, 35, 52-55] and once each in an Australian [51] and a British [21] sample. Internal consistency results are available for the linguistic validated British [46], Taiwanese [47, 56], French [48, 57], Spanish [49] and German [50] versions. In summary, the results varied between the subscales and country-specific versions of the DQol. Test-retest reliability was only assessed twice for the US [11, 55] and French versions [48, 57] and once for the British [46] and Spanish [49] versions, with heterogeneous results based on several methodological weaknesses.

#### MCQ

The MCQ is a recently developed measurement for people with mild cognitive impairment, and it is only available in its original British version [31]. The internal consistency was tested

in one methodologically strong study [31], with sufficient results; the test-retest reliability results are missing.

#### *3.4 Self- and proxy-rated Qol measurements*

##### AAIQOL

The AAIQOL allows for self- and proxy-rating of Qol in people with mild to severe dementia [32]. Only the original US version exists; there are no linguistic validated versions for other countries.

For the AAIQOL, which is also called Qol-D, internal consistency was evaluated once in one study [53], with heterogeneous results depending on self- or proxy-ratings. Test-retest [32] and inter-rater reliability [53] were also assessed once. The test-retest reliability demonstrated varying results, depending on the items for both measurement versions [32]. The inter-rater reliability of the proxy-rating version showed nearly perfect results [53]. Both studies have methodological limitations.

##### CBS Qol

The CBS Qol is a self- and proxy-rating measurement that was developed for people with mild to moderate dementia. It is available in its original US version [33] and in a linguistic validated Spanish version; the quality of the linguistic validation process was moderate [58]. Internal consistency of the CBS Qol was determined twice for the original US version [33, 59] and once for the Spanish version [58]. An evaluation of the test-retest reliability is missing, and only one assessment of inter-rater reliability for the proxy-version of the CBS Qol exists [33]. The three reliability studies showed satisfying results. However, the specific weaknesses of internal validity must be considered in each reliability study.

##### DEMQOL (self- and proxy-rated)

The dementia-specific measurements DEMQOL and DEMQOL proxy were originally developed in Britain [21] for people with mild to severe dementia. Both versions were linguistically validated to a German [60] and a Spanish version [61]. The quality of the linguistic validation for the Spanish version [61] was judged to be insufficient. The quality could not be assessed for the German version because the conference abstract is not accessible [60, 62]. Internal consistency of the British version was reported twice in one paper [21] and that of the German [60] and Spanish versions [61] was reported once. The results were good for the British and Spanish versions and heterogeneous for the German version. The test-retest reliability was investigated for the British [21] and Spanish versions [61], with satisfying results. For all results, several methodological limitations must be considered. An investigation of the inter-rater reliability of the DEMQOL proxy was not identified.

H.I.L.DE.

H.I.L.DE. is a QoL measurement for people with mild to severe dementia. It is mainly based on proxy-rating; however, for some items, self-rating is required [34, 63]. The original version is German, and no version for another country exists.

In one weak methodological study, the test-retest and inter-rater reliability were investigated and demonstrated heterogeneous results depending on the particular item [63].

QoL-AD and QoL-AD proxy

The QoL-AD (proxy- and self-rating) [9] is the most widely used dementia-specific QoL measurement for people with mild to severe dementia. For the QoL-AD, there are linguistically validated versions for Great Britain [64], Brazil [65], Taiwan [66], Japan [67], Mandarin [68] and Cantonese China [69], France [70], Spain [71], Turkey [72] and Portugal [73]. The quality

of these linguistic validations varied between good [70], moderate [65, 68, 72, 73] and insufficient [64, 67, 69, 71].

Internal consistency of the original US version was evaluated twice for the self- and proxy-rated versions [9, 74] and once for a combined self- and proxy-rated version [75]. Internal consistency was assessed twice each for the British [21, 76], Brazilian [65, 77], Japanese [67, 78], Chinese Mandarin [68, 79], French [57, 70] and Spanish versions [71, 80]. For the Taiwanese [66], Chinese Cantonese [69], Turkish [72] and Portuguese versions [73], one study was available. One study was also available for the German proxy QoI-AD version; however, no information is available concerning its linguistic validation [81]. For all investigated QoI-AD versions, the internal consistency was sufficient. Test-retest reliability was investigated twice for the US [75, 82], Chinese Mandarin [68, 79] and French versions [48, 57] and once for the British [76], Brazilian [65], Japanese [67], Chinese Cantonese [69], Spanish [80] and Portuguese versions [73]. The results showed satisfactory results for all versions, except for the Spanish QoI-AD proxy version [80]. However, the different methodological weaknesses of the reliability studies (internal consistency and test-retest reliability) must be considered. Studies on the inter-rater reliability of the QoI-AD proxy version are missing.

#### QoI-AD NH (self and proxy)

The QoI-AD nursing home version is based on the original version, which is also used for people with mild to severe dementia [9]. For the adaption to the nursing home setting, two of the original 13 items were deleted, four items were added and the wording of three items was changed [35]. In previous reviews, the two QoI-AD versions were described together. However, due to the substantial differences between the two versions, we treated them as single instruments. For the QoI-AD nursing home version, a self- and a proxy-rating version is available. In addition to the application in a US setting, the original version was applied to an

Australian sample, without linguistic validation [51]. Currently, no version for another language is available.

The internal consistency of the self- and proxy-rated versions was studied in three US [35, 52, 53] studies and one Australian study [51], with sufficient results. No studies addressing test-retest reliability were identified. The inter-rater reliability of the proxy-version was investigated in one US study [53], and it demonstrated nearly perfect agreement. To interpret the reliability properties, the methodological limitations of each study must be considered.

#### QOLAS

The QOLAS is a dementia-specific measurement for the QoI rating of people with mild to moderate dementia [36]. Beyond the original British version, no other version is available. The internal consistency of the self and proxy-QOLAS version was investigated in one study that had methodological weaknesses; this study showed a sufficient Cronbach's alpha [36]. No studies on the test-retest and inter-rater reliability of the QOLAS were identified.

#### *3.5 Self- and proxy-rated QoI measurements*

##### ADRQL

The ADRQL is a proxy-rated QoI measurement for people with mild to severe dementia. In addition to the original US [37] version, linguistic validated versions are available for Switzerland [83] and Japan [84]. The quality of both linguistic validation processes was assessed as insufficient.

The internal consistency of the original US version was investigated four times [35, 52, 53, 85] and that of the German-Swiss version was investigated once [83]. The internal consistency was tested in Germany, whereas no information was reported about the applied version and a possible linguistic validation [81]. For the Japanese version, the internal

consistency was also tested twice. In the first Japanese study, the ADRQL with 48 items and five subscales was tested and named AD-HRQL-J [84]. In the second study, the item number was reduced to 24 items that represented three subscales, *interacting with surroundings*, *expressing self* and *experiencing minimum negative behaviors*, and named QLDJ [86].

Depending on the subscales, heterogeneous results were found for internal consistency. All studies, except Yamamota et al. 2002 [86], had several methodological limitations. The test-retest reliability was only assessed for the two Japanese versions of the measurement, with varying results. The study with the highest internal validity showed good test-retest reliability [86].

Inter-rater reliability was investigated for the original US version [53], the Swiss version [83] and twice for the Japanese version [84, 86]. The results were good for the US and Japanese versions and weak for the Swiss version. The inter-rater reliability studies showed several methodological limitations.

#### CDQLP

The CDQLP, which was developed in the US, is a proxy-measurement for the assessment of QoI in people with mild to severe dementia [38]. Neither linguistic validated versions for other languages nor reliability studies exist for this measure.

#### DCM

DCM is a dementia-specific QoI measurement that is based on proxy observations [39]. The original British version was developed in several steps. The identified studies applied the 7<sup>th</sup> DCM version. In addition to two British studies [87, 88], we identified one German [89] and two US reliability studies [53, 90]. A linguistic validation was not described for the application in Germany or the US.

No study investigated the internal consistency of DCM, and only one study tested the test-retest reliability. Test-retest reliability was evaluated for the British version, and it demonstrated unsatisfying results based on a weak study design [87]. The inter-rater reliability was investigated once for the British [88] and German versions [89] and two times for the US version [53, 90]. The studies show heterogeneous results, depending on the observed items. All inter-rater reliability studies lacked internal validity.

#### OQOL and OQOLDA

The OQOLD version for mild to moderate and the OQOLDA version for severe dementia are observation-based proxy QoI measurements that were developed in the US [40]. No version for another language exists.

Only one methodologically weak reliability study investigated these measures' inter-rater reliability. This study suggested sufficient inter-rater reliability [40].

#### PWP-CIP

The PWP-CIP is a proxy-rating QoI measurement for people with mild to moderate dementia and was developed in the US [41]. No other linguistic validated version is available.

Only one study evaluated the internal consistency of this measure [41]. The study was of sufficient methodological quality and showed sufficient internal consistency.

#### QoI-D

Dementia-specific QoI measurement is a proxy-based QoI-D that was developed in Japan [42]. No version to any other language is available.

One well-designed study demonstrated the internal consistency of the original version and showed sufficient inter-rater reliability. The inter-rater reliability results are based on methodological limitations [42]. Further reliability studies do not exist.

## QUALID

The QUALID allows QoL assessments of people with moderate to severe dementia. In addition to the original US version [43], linguistic validated versions are available for Sweden [91], Spain [92] and Norway [93]. The quality of the linguistic validation ranged between insufficient [91] and moderate [92]. For the Norwegian version, all information on the linguistic validation were reported in a conference abstract that is not available electronically [93].

For each QUALID version, one investigation of internal consistency exists, demonstrating satisfactory results. The Spanish [92] and Norwegian studies [93] showed high internal validity, whereas the original US [43] and Swedish studies [91] showed multiple methodological weaknesses. Test-retest reliability was assessed for the latest two versions, demonstrating a high consistency of QoL ratings based on weak study designs [43, 91]. Inter-rater reliability was evaluated once for the US [43], Swedish [91] and Spanish versions [92], demonstrating heterogeneous results. In addition, the methodological quality of the studies was influenced by several weaknesses.

## QUALIDEM

QUALIDEM is a proxy-based measurement that consists of one version for people with mild to severe dementia and one consecutive version for people with severe dementia [10]. In addition to the original Dutch version, one linguistic validated German version is available. The quality of the employed validation process was insufficient [94].

The internal consistency was assessed twice for the Dutch version [10, 95] and three times for the German version [81, 94, 96]. For both versions, the studies demonstrated heterogeneous results depending on the subscales. The internal validity was high in one Dutch [95] and German study [96]. The other three studies showed multiple methodological

shortcomings [10, 81, 94]. Test-retest reliability was investigated once for the Dutch [10] and German versions [16], showing satisfactory results. The Dutch study had weak internal validity, whereas the German investigation showed high internal validity. The same two studies investigated inter-rater reliability, with unsatisfying results. The methodological quality was equal, as mentioned for test-retest reliability.

#### Vienna List

The Vienna List allows QoL ratings for people with severe dementia [44]. The original version was developed in Austria, and no other is available.

The internal consistency of the Vienna List was studied once and demonstrated good consistency for nurse and physician ratings. In the same study, the inter-rater reliability was evaluated, demonstrating heterogeneous results, depending on the subscales. The investigation of both reliability properties had methodological shortcomings [44]. No studies examined this measure's test-retest reliability.

#### 4. Discussion

This systematic review follows the criteria of the AMSTAR tool [97] and is the first review of dementia-specific QoL measurements to report a detailed quality appraisal for each of the included studies. Based on the systematic literature search, 19 dementia-specific QoL measurements, which are heterogeneous according to their perspectives (i.e., self- or proxy-rating) and operationalization of QoL in terms of subscales and items, were identified. Nine of these measurements are also used in countries other than the original country (i.e., DQoL, CBS QoL, DEMQOL and DEMQOL proxy, QoL-AD, QoL-AD NH, ADRQL, DCM, QUALID, QUALIDEM). The quality of the conducted linguistic validation processes for these measurements varied between insufficient and good [50, 70, 72, 92]. In six studies, dementia-specific QoL measurements were used in countries other than that of the original

measurement version but no information was available about the linguistic validation [21, 48, 51, 53, 66, 81]; two studies cited conference abstracts that are not accessible as references for the linguistic validation process [60, 93].

No linguistic validation followed all five steps recommended by Beaton et al. [25]. The steps synthesis, back-translation, expert committee and pilot testing were reported in less than half of the studies, and the quality rating of these steps varied. All adaptation processes were performed for the main language of the target country. Thus, no linguistic validation was performed for a minority group of immigrants in the target country. Beaton's linguistic validation approach was published 15 years ago, and only one study [84] that was included in the current review was published prior to that time. The large number of weak linguistic validation processes may hinder the development of equivalent measurements across countries, which are the basis for cross-national comparisons of QoL data. Moreover, the weak linguistic validations must be interpreted as a barrier to the application of dementia-specific QoL measurements in multinational research projects. Based on the increasing number of multinational research projects, this methodical limitation must be solved in the future. Therefore, the performance of linguistic validation processes based on international guidelines is strongly recommended to ensure the development of equivalent scales across countries [27]. In addition, a clear description of the applied measurement version is recommended.

Internal consistency, which is certainly the weakest reliability property, is the most frequently tested reliability property. Of the 67 reliability studies, 58 investigated the internal consistency of 15 QoL measurements. Using the Terwee et al. criteria [29], the methodological quality of only 14 internal consistency evaluations were rated as positive [31, 41, 42, 45, 55, 71, 76, 81, 86, 91-93, 95, 96]. Methodological weaknesses included small sample sizes [9, 10, 21, 30, 33, 35, 36, 43, 44, 46-49, 51-54, 56-61, 65, 66, 68, 72, 75, 77-81, 84, 94], no analysis of the factor structure of the measurement [9, 11, 21, 30, 35, 36, 46-54,

56-60, 65-70, 72-75, 77-80, 83-85], missing information on the characteristics of the study participants and missing information on the subsample for evaluation of internal consistency [10, 53], missing information on characteristics of the proxy-raters [33, 35, 43, 44, 52, 53, 58, 84-86, 93-96] and unreliable calculation of the statistical measurements [21, 52, 87].

A total of 31 studies reported test-retest reliability results for 10 measurements, and 20 studies reported inter-rater reliability for 11 measurements. The quality rating of these studies resulted in a positive rating for 1 test-retest reliability [16] and 1 inter-rater reliability study [92] that used the QUAREL tool [28]. The internal validity of test-retest and inter-rater reliability studies was threatened by small sample sizes [11, 21, 30, 32, 33, 40, 42-46, 49, 53, 61, 63, 65, 68, 69, 73, 75, 76, 79, 80, 82-84, 86-89, 91], missing information on the characteristics of study participants [10, 32, 42-44, 53, 63, 65, 67-69, 73, 79, 80, 82, 83, 88, 90, 91] or proxy-raters [10, 32, 33, 40, 42-44, 53, 63, 65, 67-69, 73, 79, 80, 82-84, 86-88, 90, 91], inappropriate statistical measurements [11, 55, 75], missing determination of statistical uncertainty (e.g., confidence intervals [CIs]) [10, 11, 21, 28, 32, 33, 40, 42, 44, 46, 48, 49, 53, 57, 63, 65, 68-70, 76, 79, 82-84, 86-88, 90, 91], and missing information on blinding between the first and second observations or different proxy-raters [10, 11, 21, 32, 33, 40, 42, 44-46, 48, 49, 53, 55, 57, 61, 63, 65, 67-70, 73, 76, 79, 80, 82-84, 86, 87, 91]. Because reliability is a necessary condition of the validity of a measurement, these results are surprising and highlight the need for methodologically well-designed reliability studies in the field of dementia-specific QoL measurements. An in-depth analysis of the test-retest and inter-rater reliabilities of different measurement items to gather a more detailed understanding of the properties of more or less reliable items was not feasible because of the missing information on the reported results at the item level.

The need for reliable dementia-specific QoL measurements is also illustrated by the need for effective psycho-social or non-pharmacological interventions in the field of dementia research and practice. The effectiveness of numerous interventions has been investigated in

the past, and these studies demonstrated mostly unsatisfactory results [98]. One possible explanation for these results is insufficient knowledge about the reliability properties of applied dementia-specific QoL measurements, with negative consequences for the validity of the measures' QoL values.

Based on the linguistic validation and reliability results, the self-rated measurements DQoL and QoL-AD and the proxy-measurements QUALID and QUALIDEM seem to be the most widely studied measurements. However, the results of the current review demonstrate that these measurements also require further investigation. In general, prior to the application of a measurement in observational or experimental studies and in accordance with Schölzel-Dorenbos et al. [20], the researchers must select the most appropriate measurement based on the measurement perspective, the recommended dementia severity stage and setting for the particular measurement, and the operationalization of the concept of QoL. In a second step, the investigation of the (likely) missing linguistic validation steps and reliability properties is recommended. To realize the needed linguistic validation steps and to evaluate the psychometric properties, financial funding must be increased for of these methodological steps.

#### 4.1 Strengths and limitations

This review is the first to consider the quality of the linguistic validation process. Thus, this review adds to the body of knowledge on dementia-specific QoL measurements and delivers a more detailed analysis compared to earlier reviews [7, 8, 17-22]. As a limitation, note that according to the inclusion and exclusion criteria, only the studies that addressed measurement development and psychometric testing were included. Other designs that may have provided information regarding the measurement properties were excluded. English and German language studies were considered for inclusion. Therefore, a possible language bias is not excluded. Because QoL is one major outcome in interventions, future studies on

the psychometric properties of QoI measurements should be strengthened as a research priority. A financially independent international database that aims to register QoI measurement applications and that generates results on psychometric properties is desirable. Such a database will allow the estimation of possible language biases and decrease the possibility of publication biases in future reviews.

According to its objective, this review is limited to the linguistic validation and reliability of measurement scales. For a detailed evaluation of the validity of these measurements, a second review will be published elsewhere. As mentioned in the review by Bowling et al. (2015), a rigorous appraisal scheme must be developed for a detailed validity review, which will allow the appraisal of the relationship of the conceptual framework of each QoI measurement and the applied validation strategies [8].

#### 4.2 Conclusions

In summary, it has become clear that more research is needed to establish strong dementia-specific QoI measurements, which could be used for the evaluation of interventions in the context of dementia care or health care services. In addition to reliability studies, this review shows a need for rigorous linguistically validated measurements. Without such measurements, the use of the same dementia-specific QoI measurements in different countries is questionable. For the performance of linguistic validations and reliability studies of dementia-specific QoI measurements, the application of international guidelines and quality criteria is strongly recommended. In addition, the analysis and reporting of reliability properties for each measurement item is recommended to allow an in-depth analysis of the properties of more or less reliable items.

#### 5. Description of the authors' roles

Study design: MND, CGGS, GM, SB, MH

Data analysis: MND, CGGS

Manuscript preparation: MND, CGGS; GM, SB, MH

ACCEPTED MANUSCRIPT

**References**

- [1] ADI. Dementia statistics. 2009 [07.01.2015]; Available from: <http://www.alz.co.uk/research/statistics>.
- [2] Moniz-Cook E, Vernooij-Dassen M, Woods R, Verhey F, Chattat R, De Vugt M, et al. A European consensus on outcome measures for psychosocial intervention research in dementia care. *Aging Ment Health* 2008;12(1):14-29.
- [3] Rabins PV, Black BS. Measuring quality of life in dementia: purposes, goals, challenges and progress. *Int Psychogeriatr* 2007;19(3):401-7.
- [4] Lawton MP. Quality of life in Alzheimer disease. *Alzheimer Dis Assoc Disord* 1994;8 Suppl 3:138-50.
- [5] Jonker C, Gerritsen DL, Bosboom PR, Van der Steen JT. A model for quality of life measures in patients with dementia: Lawton's next step. *Dement Geriatr Cogn Disord* 2004;18(2):159-64.
- [6] Ettema TP, Droes RM, de Lange J, Ooms ME, Mellenbergh GJ, Ribbe MW. The concept of quality of life in dementia in the different stages of the disease. *Int Psychogeriatr* 2005;17(3):353-70.
- [7] Perales J, Cosco TD, Stephan BC, Haro JM, Brayne C. Health-related quality-of-life instruments for Alzheimer's disease and mixed dementia. *Int Psychogeriatrics* 2013;25(5):691-706.
- [8] Bowling A, Rowe G, Adams S, Sands P, Samsi K, Crane M, et al. Quality of life in dementia: a systematically conducted narrative review of dementia-specific measurement scales. *Aging Ment Health* 2015;19(1):13-31.
- [9] Logsdon RG, Gibbons LE, McCurry SM, Teri L. Quality of life in Alzheimer's Disease: Patient and caregiver reports. *J Ment Health Aging* 1999;5(1):21-32.
- [10] Ettema TP, Droes 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):549-56.
- [11] Brod M, Stewart AL, Sands L, Walton P. Conceptualization and measurement of quality of life in dementia: the dementia quality of life instrument (DQoL). *Gerontologist* 1999;39(1):25-35.
- [12] Gräske J, Fischer T, Kuhlmeier A, Wolf-Ostermann K. Quality of life in dementia care - differences in quality of life measurements performed by residents with dementia and by nursing staff. *Aging Ment Health* 2012;16(7):819-27.
- [13] Sands LP, Ferreira P, Stewart AL, Brod M, Yaffe K. What explains differences between dementia patients' and their caregivers' ratings of patients' quality of life? *Am J Geriatr Psychiatry* 2004;12(3):272-80.
- [14] Huang HL, Chang MY, Tang JS, Chiu YC, Weng LC. Determinants of the discrepancy in patient- and caregiver-rated quality of life for persons with dementia. *J Clin Nurs* 2009;18(22):3107-18.
- [15] Gräske J, Meyer S, Wolf-Ostermann K. Quality of life ratings in dementia care - a cross-sectional study to identify factors associated with proxy-ratings. *Health Qual Life Outcomes* 2014;12(1):177.
- [16] Dichter M, Schwab C, Meyer G, Bartholomeyczik S, Dortmann O, Halek M. Measuring the quality of life in mild to very severe dementia: Testing the inter-rater and intra-rater reliability of the German version of the QUALIDEM. *International Psychogeriatrics* 2014;doi:10.1017/S1041610214000052.
- [17] Ettema TP, Droes RM, de Lange J, Mellenbergh GJ, Ribbe MW. A review of quality of life instruments used in dementia. *Qual Life Res* 2005;14(3):675-86.

- [18] Oppikofer S. Quality of life in dementia. Identification, review and documentation of existing instruments to measure quality of life in people with severe dementia. [Lebensqualität bei Demenz. Eine Bestandsaufnahme, Sichtung und Dokumentation bestehender Instrumente zur Messung der Lebensqualität bei Menschen mit schwerer Demenz]. Zürich: University Hospital Zürich 2008.
- [19] Ready RE, Ott BR. Quality of Life measures for dementia. *Health Qual Life Outcomes* 2003;1:11.
- [20] Scholzel-Dorenbos CJ, Ettema TP, Bos J, Boelens-van der Knoop E, Gerritsen DL, Hoogeveen F, et al. Evaluating the outcome of interventions on quality of life in dementia: selection of the appropriate scale. *Int J Geriatr Psychiatry* 2007;22(6):511-9.
- [21] Smith SC, Lamping DL, Banerjee S, Harwood RH, Foley B, Smith P, 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 Technology Assessment NHS R&D HTA Programme* 2005;9(10):1-110.
- [22] Walker MD, Salek SS, Bayer AJ. A review of quality of life in Alzheimer's disease Part 1: issues in assessing disease impact. *PharmacoEconomics* 1998;14:499-530.
- [23] Graske J, Fischer T, Kuhlmeier A, Wolf-Ostermann K. Dementia-specific quality of life instruments and their appropriateness in shared-housing arrangements--a literature study. *Geriatric nursing* 2012;33(3):204-16.
- [24] Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009;62(10):1013-20.
- [25] Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000;25(24):3186-91.
- [26] Acquadro C, Conway K, Hareendran A, Aaronson N, European Regulatory I, Quality of Life Assessment G. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value Health* 2008;11(3):509-21.
- [27] Uysal-Bozkir O, Parlevliet JL, de Rooij SE. Insufficient cross-cultural adaptations and psychometric properties for many translated health assessment scales: a systematic review. *J Clin Epidemiol* 2013;66(6):608-18.
- [28] Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol* 2010;63(8):854-61.
- [29] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34-42.
- [30] Trigg R, Jones RW, Skevington SM. Can people with mild to moderate dementia provide reliable answers about their quality of life? *Age Ageing* 2007;36(6):663-9.
- [31] Dean K, Jenkinson C, Wilcock G, Walker Z. The development and validation of a patient-reported quality of life measure for people with mild cognitive impairment. *Int Psychogeriatr* 2014;26(3):487-97.
- [32] Albert SM, Del Castillo-Castaneda C, Sano M, Jacobs DM, Marder K, Bell K, et al. Quality of life in patients with Alzheimer's disease as reported by patient proxies. *J Am Geriatr Soc* 1996;44(11):1342-7.
- [33] Ready RE, Ott BR, Grace J, Fernandez I. The Cornell-Brown Scale for Quality of Life in dementia. *Alzheimer Dis Assoc Disord* 2002;16(2):109-15.

- [34] Becker S, Kruse A, Schroder J, Seidl U. [The Heidelberg instrument for the assessment of quality of life in dementia (H. I. L. DE.) - dimensions of quality of life and methods of organization]. *Z Gerontol Geriatr* 2005;38(2):108-21.
- [35] Edelman P, Fulton BR, Kuhn D, Chang CH. A comparison of three methods of measuring dementia-specific quality of life: perspectives of residents, staff, and observers. *Gerontologist* 2005;45 Spec No 1(1):27-36.
- [36] Selai C, Trimble MR, Rossor MN, Harvey RJ. Assessing quality of life in dementia: preliminary psychometric testing of the Quality of Life Assessment Schedule (QOLAS). *Neuropsychological Rehabilitation* 2001;11:219-43.
- [37] Rabins PV, Kasper JD, Kleinman L, Black BS, Patrick DL. Concepts and methods in the development of the ADRQL: An instrument for assessing health-related quality of life in persons with Alzheimer's disease. *J Ment Health Aging* 1999;5:33-48.
- [38] Salek M, Schwartzberg A, Bayer A. Evaluating health-related quality of life in patients with dementia: development of a proxy self-administered questionnaire. *Pharmacy World and Science* 1996;18(A6).
- [39] Kitwood T, Bredin K. A new approach to the evaluation of dementia care. *Journal of Advances in Health and Nursing Care* 1992;1(5):41-60.
- [40] Edelman P, Fulton BR, Kuhn D, Gallager M, Dougherty J, Long C. Assessing Quality of Life Across the Dementia Continuum. Two New Observational Tools for Researchers and Practitioners. *Azheimer's Care Today* 2007;8(4):332-43.
- [41] Burgener SC, Twigg P, Popovich A. Measuring psychological well-being in cognitively impaired persons. *Dementia* 2005;4(4):463-85.
- [42] Terada S, Ishizu H, Fujisawa Y, Fujita D, Yokota O, Nakashima H, et al. Development and evaluation of a health-related quality of life questionnaire for the elderly with dementia in Japan. *Int J Geriatr Psychiatry* 2002;17(9):851-8.
- [43] 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):114-6.
- [44] Porzolt F, Kojer M, Schmidl M, Greimel ER, Sigle J, Richter J, et al. A new instrument to describe indicators of well-being in old-old patients with severe dementia - the Vienna List. *Health Qual Life Outcomes* 2004;2:10.
- [45] Trigg R, Skevington SM, Jones RW. How can we best assess the quality of life of people with dementia? The Bath Assessment of Subjective Quality of Life in Dementia (BASQID). *Gerontologist* 2007;47(6):789-97.
- [46] Karim S, Ramanna G, Petit T, Doward L, Burns A. Development of the Dementia Quality of Life questionnaire (D-QOL): UK version. *Aging Ment Health* 2008;12(1):144-8.
- [47] Chiu YC, Chiu YF, Hsu WC, Lee SH, Chen ST, Kao HY. Developing a Chinese quality of life in dementia instrument for patients with early-to-moderate dementia: an exploratory test of validity. *J Clin Nurs* 2010;19(15-16):2174-84.
- [48] Wolak A, Jolly D, Dramé M, Boyer F, Morrone I, Aquino JP, et al. Quality of life in dementia: Psychometric properties of a French language version of the Dementia Quality of Life questionnaire (DQoL). *European Geriatric Medicine* 2010;1(6):334-47.
- [49] Lucas-Carrasco R, Gomez-Benito J, Rejas J, Brod M. The Spanish version of the dementia quality of life questionnaire: a validation study. *Aging Ment Health* 2011;15(4):482-9.

- [50] Voigt-Radloff S, Leonhart R, Schutzwahl M, Jurjanz L, Reuster T, Gerner A, et al. Dementia quality of life instrument--construct and concurrent validity in patients with mild to moderate dementia. *Eur J Neurol* 2012;19(3):376-84.
- [51] Moyle W, Murfield JE, Griffiths SG, Venturato L. Assessing quality of life of older people with dementia: a comparison of quantitative self-report and proxy accounts. *J Clin Nurs* 2012;21(11-12):1632-40.
- [52] Edelman P, Fulton BR, Kuhn D. Comparison of dementia-specific quality of life measures in adult day centers. *Home health care services quarterly* 2004;23(1):25-42.
- [53] Sloane PD, Zimmerman S, Williams CS, Reed PS, Gill KS, Preisser JS. Evaluating the quality of life of long-term care residents with dementia. *Gerontologist* 2005;45 Spec No 1(1):37-49.
- [54] Ready RE, Ott BR, Grace J. Insight and cognitive impairment: effects on quality-of-life reports from mild cognitive impairment and Alzheimer's disease patients. *Am J Alzheimers Dis Other Demen* 2006;21(4):242-8.
- [55] Adler E, Resnick B. Reliability and validity of the Dementia Quality of Life measure in nursing home residents. *Western journal of nursing research* 2010;32(5):686-704.
- [56] Chiu YC, Shyu Y, Liang J, Huang HL. Measure of quality of life for Taiwanese persons with early to moderate dementia and related factors. *Int J Geriatr Psychiatry* 2008;23(6):580-5.
- [57] Wolak-Thierry A, Novella JL, Barbe C, Morrone I, Mahmoudi R, Jolly D. Comparison of QoL-AD and DQoL in elderly with Alzheimer's disease. *Aging Ment Health* 2014:1-5.
- [58] Lucas-Carrasco R, Gomez-Benito J, Rejas J, Ott BR. The Cornell-Brown scale for quality of life in dementia: Spanish adaptation and validation. *Alzheimer Dis Assoc Disord* 2013;27(1):44-50.
- [59] Ready RE, Ott BR. Integrating Patient and Informant Reports on the Cornell-Brown Quality-of-Life Scale. *American Journal of Alzheimer's Disease & Other Dementias* 2008;22(6):528-34.
- [60] Berwig M, Leicht H, Gertz HJ. Critical evaluation of self-rated quality of life in mild cognitive impairment and Alzheimer's disease - further evidence for the impact of anosognosia and global cognitive impairment. *J Nutr Health Aging* 2009;13(3):226-30.
- [61] Lucas-Carrasco R, Lamping DL, Banerjee S, Rejas J, Smith SC, Gomez-Benito J. Validation of the Spanish version of the DEMQOL system. *Int Psychogeriatr* 2010;22(4):589-97.
- [62] Knorr C, Fröhlich L. What determines health related quality of life (HRQOL) in patients with dementia and their caregivers? 2007. *International Society for Quality of Life Research meeting abstracts* 2007.
- [63] Becker S, Kaspar R, Kruse A. [The Heidelberg instrument for the assessment of quality of life in dementia (H. I. L. DE.)]. Bern: Hans Huber; 2011.
- [64] Selai C, Vaughan A, Harvey RJ, Logsdon R. Using the QOL-AD in the UK. *Int J Geriatr Psychiatry* 2001;16(5):537-8.
- [65] Novelli MM, Dal Rovere HH, Nitrini R, Caramelli P. Cross-cultural adaptation of the quality of life assessment scale on Alzheimer disease. *Arquivos de neuro-psiquiatria* 2005;63(2A):201-6.
- [66] Fuh JL, Wang SJ. Assessing quality of life in Taiwanese patients with Alzheimer's disease. *Int J Geriatr Psychiatry* 2006;21(2):103-7.
- [67] Matsui T, Nakaaki S, Murata Y, Sato J, Shinagawa Y, Tatsumi H, et al. Determinants of the quality of life in Alzheimer's disease patients as assessed by the

- Japanese version of the Quality of Life-Alzheimer's disease scale. *Dement Geriatr Cogn Disord* 2006;21(3):182-91.
- [68] Lin Kiat Yap P, Yen Ni Goh J, Henderson LM, Min Han P, Shin Ong K, Si Ling Kwek S, et al. How do Chinese patients with dementia rate their own quality of life? *Int Psychogeriatr* 2008;20(3):482-93.
- [69] Chan I, Chu L, Lee P, Li S, Yu K. Effects of cognitive function and depressive mood on the quality of life in Chinese Alzheimer's disease patients in Hong Kong. *Geriatr Gerontol Int* 2011;11:69-76.
- [70] Wolak A, Novella JL, Drame M, Guillemin F, Di Pollina L, Ankri J, et al. Transcultural adaptation and psychometric validation of a French-language version of the QoL-AD. *Aging Ment Health* 2009;13(4):593-600.
- [71] Gomez-Gallego M, Gomez-Garcia J, Ato-Garcia M. Confirmatory factor analysis of the quality of life in Alzheimer's disease scale in patients with Alzheimer's disease. *Experimental aging research* 2014;40(3):266-79.
- [72] Akpınar B, Küçükgülü Ö. The Validity and Reliability of the Turkish version of the Quality of life scale for patients with Alzheimer's Disease (QoL-AD). *Journal of Neurological Sciences* 2012;29(3):554-65.
- [73] Barrios H, Verdelho A, Narciso S, Goncalves-Pereira M, Logsdon R, de Mendonca A. Quality of life in patients with cognitive impairment: validation of the Quality of Life-Alzheimer's Disease scale in Portugal. *Int Psychogeriatr* 2013;25(7):1085-96.
- [74] Logsdon RG, Gibbons LE, McCurry SM, Teri L. Assessing quality of life in older adults with cognitive impairment. *Psychosom Med* 2002;64(3):510-9.
- [75] Hoe J, Katona C, Roch B, Livingston G. Use of the QOL-AD for measuring quality of life in people with severe dementia - the LASER-AD study. *Age Ageing* 2005;34(2):130-5.
- [76] Thorgrimsen L, Selwood A, Spector A, Royan L, de Madariaga Lopez M, Woods RT, et al. Whose quality of life is it anyway? The validity and reliability of the Quality of Life-Alzheimer's Disease (QoL-AD) scale. *Alzheimer Dis Assoc Disord* 2003;17(4):201-8.
- [77] Novelli MM, Nitrini R, Caramelli P. Validation of the Brazilian version of the quality of life scale for patients with Alzheimer's disease and their caregivers (QOL-AD). *Aging Ment Health* 2010;14(5):624-31.
- [78] Tatsumi H, Yamamoto M, Nakaaki S, Hadano K, Narumoto J. Utility of the Quality of Life-Alzheimer's Disease Scale for mild cognitive impairment. *Psychiatry and clinical neurosciences* 2011;65(5):533.
- [79] Yu HM, He RL, Ai YM, Liang RF, Zhou LY. Reliability and validity of the quality of life-Alzheimer disease Chinese version. *J Geriatr Psychiatry Neurol* 2013;26(4):230-6.
- [80] Leon-Salas B, Logsdon RG, Olazarán J, Martínez-Martin P, The M-A. Psychometric properties of the Spanish QoL-AD with institutionalized dementia patients and their family caregivers in Spain. *Aging Ment Health* 2011;15(6):775-83.
- [81] Grasko J, Verbeek H, Gellert P, Fischer T, Kuhlmeier A, Wolf-Ostermann K. How to measure quality of life in shared-housing arrangements? A comparison of dementia-specific instruments. *Qual Life Res* 2014;23(2):549-59.
- [82] Logsdon RG, Albert SM. Assessing Quality of Life in Alzheimer's Disease: Conceptual and Methodological Issues. *J Ment Health Aging* 1999;5(1):3-6.
- [83] Menzi-Kuhn C. [Quality of life of people with dementia in institutional care] Lebensqualität von Menschen mit Demenz in stationären Langzeitpflegeeinrichtungen [Unpublished Masterthesis]. Maastricht: Maastricht University; 2006.

- [84] Yamamoto N, Abe T, Yamada Y, Yamazato C, Amemiya H, Sugishita C, et al. Reliability and validity of a Japanese quality of life scale for the elderly with dementia. *Nursing and Health Sciences* 2000;2:69-78.
- [85] Kasper JD, Black BS, Shore AD, Rabins PV. Evaluation of the validity and reliability of the Alzheimer Disease-related Quality of Life Assessment Instrument. *Alzheimer Dis Assoc Disord* 2009;23(3):275-84.
- [86] Yamamoto-Mitani N, Abe T, Okita Y, Hayashi K, Sugishita C, Kamata K. Development of a Japanese quality of life instrument for older adults experiencing dementia (QLDJ). *International journal of aging & human development* 2002;55(1):71-95.
- [87] Fossey J, Lee L, Ballard C. Dementia Care Mapping as a research tool for measuring quality of life in care settings: psychometric properties. *Int J Geriatr Psychiatry* 2002;17(11):1064-70.
- [88] Thornton A, Hatton C, Tatham A. Dementia Care Mapping reconsidered: exploring the reliability and validity of the observational tool. *Int J Geriatr Psychiatry* 2004;19(8):718-26.
- [89] Rusing D. [Inter-rater reliability of behavior and wellbeing codes of the observational tool Dementia Care Mapping (DCM)] Die Interraterreliabilität des Verhaltens- und Wohlbefindlichkeitskodierung des Beobachtungsinstrumentes „Dementia Care Mapping (DCM)“: Witten/Herdecke University; 2004.
- [90] Sloane PD, Brooker D, Cohen L, Douglass C, Edelman P, Fulton BR, et al. Dementia care mapping as a research tool. *Int J Geriatr Psychiatry* 2007;22(6):580-9.
- [91] Falk H, Persson LO, Wijk H. A psychometric evaluation of a Swedish version of the Quality of Life in Late-Stage Dementia (QUALID) scale. *Int Psychogeriatrics* 2007;19(6):1040-50.
- [92] Garre-Olmo J, Planas-Pujol X, Lopez-Pousa S, Weiner MF, Turon-Estrada A, Juvinya D, et al. Cross-cultural adaptation and psychometric validation of a Spanish version of the Quality of Life in Late-Stage Dementia Scale. *Qual Life Res* 2010;19(3):445-53.
- [93] Mjorud M, Kirkevold M, Rosvik J, Engedal K. Principal component analysis of the Norwegian version of the quality of life in late-stage dementia scale. *Dement Geriatr Cogn Disord* 2014;37(5-6):265-75.
- [94] Dichter M, Bartholomeyczik S, Nordheim J, Achterberg W, Halek M. Validity, reliability, and feasibility of a quality of life questionnaire for people with dementia. *Z Gerontol Geriatr* 2011;44(6):405-10.
- [95] Bouman AI, Ettema TP, Wetzels RB, van Beek AP, de Lange J, Droes RM. 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. *Int J Geriatr Psychiatry* 2011;26(7):711-22.
- [96] Dichter MN, Dortmann O, Halek M, Meyer G, Holle D, Nordheim J, et al. 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 Qual Life Outcomes* 2013;11:91.
- [97] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
- [98] Cooper C, Mukadam N, Katona C, Lyketsos CG, Ames D, Rabins P, et al. Systematic review of the effectiveness of non-pharmacological interventions to improve quality of life of people with dementia. *Int Psychogeriatr* 2012;24(6):856-70.

[99] Røen I, Selbaek G, Lerdal A. The reliability and validity of the Norwegian version of the Quality of life in Late-stage Dementia (QUALID) scale. International psychogeriatric association Montreal, Canada2009.

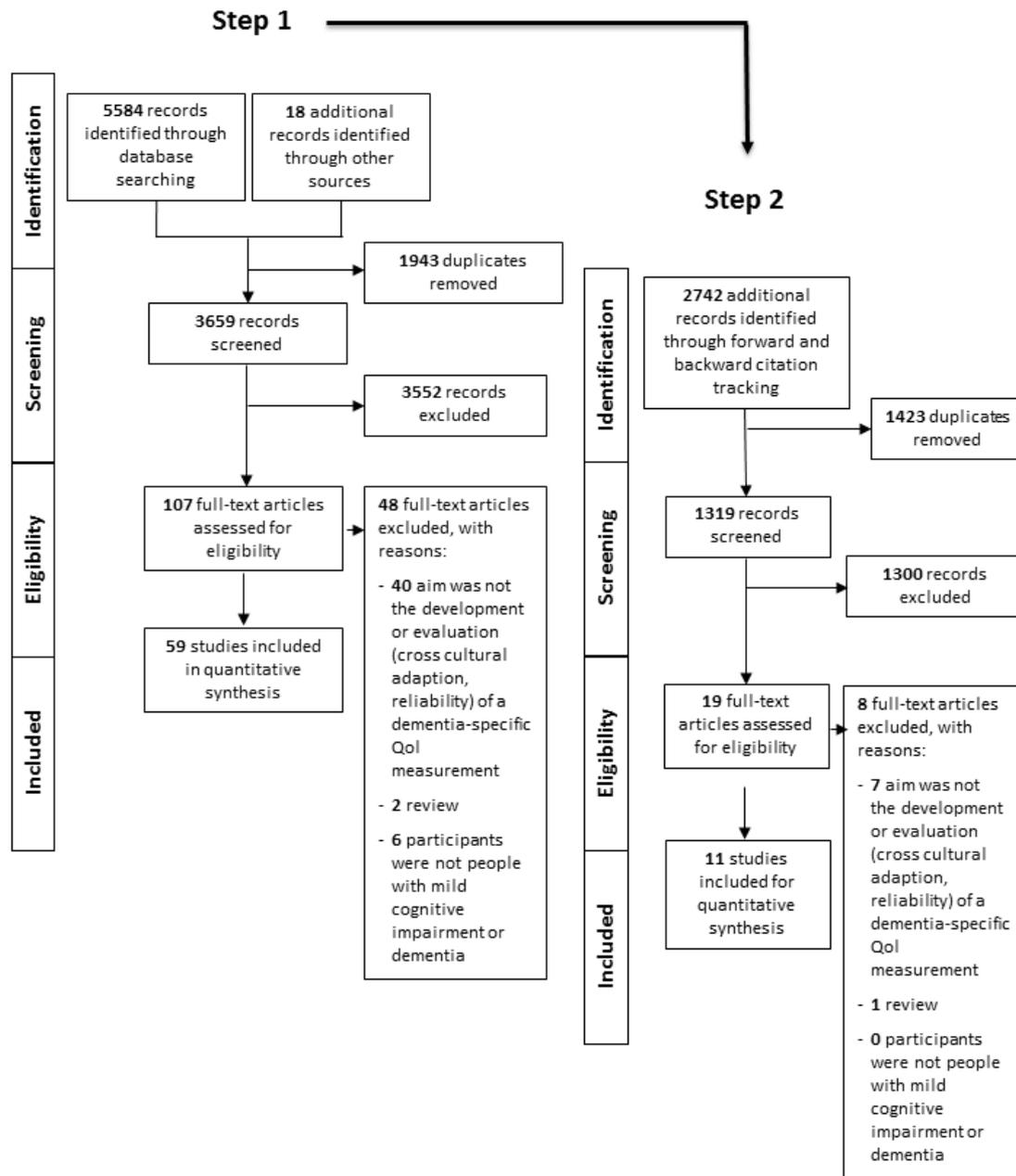
ACCEPTED MANUSCRIPT

**Table 1: Reviews on dementia-specific Qol measurements**

Review/Measurements	Walker et al. 1998 [22]	Ready et al. 2003 [19]	Ettema et al. 2005 [17]	Smith et al. 2005 [21]	Schölzel- Dorenbos et al. 2007 [20]	Oppikhofer 2008 [18]	Gräske et al. [23]	Perales et al. 2013 [7]	Bowling et al. 2014 [8]
<b>Self-rated only</b>									
BASQID							✓	✓	✓
DQol	✓	✓	✓	✓	✓	✓	✓	✓	✓
MCQ									
<b>Self-rating and proxy rating</b>									
AAIQOL	✓	✓		✓		✓	✓	✓	✓
CBS Qol		✓	✓		✓	✓	✓	✓	✓
DEMQOL & DEMQOL-proxy				✓	✓	✓	✓	✓	✓
H.I.L.DE.						✓	✓		
Qol-AD	✓	✓	✓	✓	✓	✓	✓	✓	✓
Qol-AD NH									
QOLAS		✓		✓		✓	✓	✓	✓
<b>Proxy-rating only</b>									
ADRQL & AD-HRQL-J (QLDJ)		✓	✓	✓	✓	✓	✓	✓	✓
CDQLP	✓			✓			✓	✓	
DCM	✓	✓	✓			✓	✓	✓	✓
OQOLD & OQOLDA							✓		
PWP-CIP							✓		
Qol-D			✓	✓		✓	✓	✓	✓
QUALID		✓				✓	✓	✓	✓
QUALIDEM					✓		✓	✓	✓
Vienna List					✓		✓	✓	
<b>Methodological steps</b>									
Literature search	✓	✓	✓			✓	✓	✓	✓
Data extraction									✓
Quality appraisal									
Data synthesis/summary	✓	✓	✓	✓	✓	✓	✓	✓	✓

BASQID: Bath Assessment of Subjective Quality of Life in Dementia [30], DQol: Dementia Quality of life [11], MCQ: Mild Cognitive Impairment Questionnaire [31], AAIQOL: Activity and Affect Indicator Quality of life [32], CBS Qol: Cornell-Brown Scale for Quality of Life in Dementia [33], H.I.L.DE.: Heidelberg instrument for the assessment of quality of life in dementia [34], Qol-AD: Quality of Life in Alzheimer's Disease [9], Qol-AD NH: Quality of Life in Alzheimer's Disease Nursing home version [35], QOLAS: Quality of Life Assessment Schedule [36], ADRQL: Alzheimer Disease Related Quality of Life [37] (Japanese version called AD-HRQL-J or later QLDJ) [84], CDQLP: Community Dementia Quality of Life Profile [38], DCM: Dementia Care Mapping [39], OQOLD: Observing Quality of Life in Dementia [40], OQOLDA: Observed Quality of Life in Dementia Advanced [40], PWP-CIP: Psychological Well-being in Cognitively Impaired Persons [41], Qol-D: Quality of Life for Dementia [42], QUALID: Quality of Life in Late-stage Dementia Scale [43], QUALIDEM [10], Vienna List [44]

Figure 1: A flow diagram of the literature search and study selection



(Some studies include more than one dementia-specific QoI measurement)

**Table 2: Results of the linguistic validation and reliability evaluation of dementia-specific QoL measurements**

Method of measurement / Measurements <sup>s</sup>	Version - year of publication or validation	Quality of linguistic validation	Quality of reliability evaluation		
		Overall conclusion	IC	TRR	IRR
<b>Self-rating only</b>					
BASQID	(original) GB - 2007	---	+	?	---
DQoL	(original) US - 1999	---	+	?	---
	GB - 2008	I	?	?	---
	TW - 2010	I	?	0	---
	FR - 2010	?	?	?	---
	ES - 2011	I	?	-	---
	AU - 2012	?	?	0	---
	DE - 2012	G	?	0	---
MCQ	(original) GB - 2014	---	+	0	---
<b>Self-rating and proxy-rating</b>					
AAIQOL (proxy)	(original) US - 1996	---	? (?)	0 (?)	(?)
CBS QoL (proxy)	(original) US - 2002	---	? (?)	0 (0)	(?)
	ES - 2013	M	? (0)	0 (0)	(0)
DEMQOL (proxy)	(original) GB - 2005	---	? (?)	? (?)	(0)
	DE - 2009	?	? (?)	0 (0)	(0)
	ES - 2010	I	? (?)	? (0)	(0)
H.I.L.DE.	(original) DE - 2005	---	0 (0)	? (?)	?
QoL-AD (proxy)	(original) US - 1999	---	? (?)	? (?)	(0)
	GB - 2001	I	+ (?)	? (0)	(0)
	BR - 2005	M	? (?)	? (?)	(0)
	TW - 2006	?	? (?)	0 (0)	(0)
	JP - 2006	I	? (?)	? (?)	(0)
	(Mandarin) CN - 2008	M	? (?)	? (?)	(0)
	FR - 2009	G	? (?)	? (?)	(0)
	(Cantonese) CN - 2011	I	? (?)	? (?)	(0)
	TR - 2012	M	? (?)	0 (0)	(0)
	PT - 2013	M	? (?)	? (?)	(0)
	ES - 2014	I	+ (?)	? (?)	(0)
	DE - 2014	?	0 (+)	0 (0)	(0)
	QoL-AD NH (proxy)	(original) US - 2005	---	? (?)	0 (0)
AU - 2012		?	? (0)	0 (0)	(0)
QOLAS (proxy)	(original) GB - 2001	---	? (?)	0 (0)	(0)
<b>Proxy-rated only</b>					
ADRQL /	(original) US - 1999	---	?	0	?
Japanese version called	JP - 2000	I	? / +	? / ?	? / ?
AD-HRQL-J or later QLDJ	CH - 2006	I	-	0	?
	DE - 2013	?	?	0	0
CDQLP	(original) US - 1996	---	0	0	0
DCM	(original) GB - 1992	---	0	?	-
	DE - 2004	I	0	0	-
	US - 2005	I	0	0	?
OQOLD / OQOLDA	(original) US - 2007	---	0 / 0	0 / 0	0 / ?
PWP-CIP	(original) US - 2005	---	+	0	0
QoL-D	(original) JP - 2002	---	+	0	?
QUALID	(original) US - 2000	---	?	?	?
	SE - 2007	I	+	?	-
	ES - 2010	M	+	0	+
	NO - 2014	?	+	0	0
QUALIDEM	(original) NL - 2007	---	+	?	-
	DE - 2011	I	+	+	-
Vienna List	(original) AT - 2004	---	?	0	-

<sup>a</sup> See table1 for the full names of the measurements.

IC = Internal Consistency, TRR = Test-Retest Reliability, IRR = Inter-Rater Reliability

Linguistic validation overall conclusion: G = Good; M = Moderate; I = Insufficient, ? = no information reported or cited source not available.

Reliability evaluation overall rating: + = positive rating, - = negative rating, 0 = no information available, ? = unclear.

## Appendix A

### Characteristics of dementia-specific QoL measurements

Measurement	Dementia severity	Time frame	QoL Domains/Subscales	No. of Items	Response options	Scoring procedure	Primary papers (development)	Secondary papers (modification & validation in other countries)
<b>Self-rated only</b>								
Bath Assessment of Subjective Quality of Life in Dementia, BASQID	Mild – moderate	Present	A: Life satisfaction B: Feeling of positive QoL	14 + 3 additional Item	5 point (not at all satisfied – extremely satisfied) and (not at all – a great deal) 4 point (very poor – very good) for additional Items	Total and subscales	[30]	[45]
Dementia Quality of life, DQoL	Mild – moderate	Present	A: Self-esteem B: Positive affect/humor C: Negative affect D: Feelings of belonging E: Sense of aesthetics	29 + 1 additional item rating overall QoL + 3 screening items	5 point (never – very often) and (not at all – very)	Subscales	[11]	[21, 35, 46-57]
Mild Cognitive Impairment Questionnaire, MCQ	Mild Cognitive Impairment	Present	A: Emotional effects B: Practical concerns	13 Items	5 point (never – always)	Subscales	[31]	
<b>Self-rating and proxy rating</b>								
Activity and Affect Indicator Quality of life, AAIQOL	Mild – severe	Previous week (activity) respectively two (affect) weeks	A: Affect B: Activity	21	Affect: 9 point (never – several times per day) Activity: Frequency (3 point = ≤ three times per week, occasionally ≥ three times per week, never), Enjoyment (2 point = Yes/No)	Subscales	[32]	[53]
Cornell-Brown Scale for Quality of Life in Dementia, CBS QoL	Mild – moderate	Previous month	▪ None subscales ▪ Items represent the QoL domains positive affect, negative affect, satisfaction, physical complaints	19	5 point (-2 = severe or constant presence of the negative end of an item – +2 = very or constant presence of the positive end of an item)	Total	[33]	[58, 59]
DEMQL (proxy)	Mild – moderate (Mild – severe)	Previous week	A: Daily activities B: Memory C: Negative emotion D: Positive emotion A (proxy): Functioning	28 (31) + 1 additional Item	4 point (a lot – not at all)	Total	[21]	[60, 61]

Measurement	Dementia severity	Time frame	Qol Domains/Subscales	No. of Items	Response options	Scoring procedure	Primary papers (development)	Secondary papers (modification & validation in other countries)
B (proxy): Emotion								
Heidelberg instrument for the assessment of quality of life in dementia, H.I.L.DE.	Mild – severe	Combination of previous and special observations	A: Medical care and pain experience B: Physical environment C: Activities D: Social reference system E: Emotional-being and life satisfaction K: Patterns of competence	48	varying response options per item	Subscales	[34]	[63]
Quality of Life in Alzheimer's Disease, QoL-AD (proxy)	Mild – severe	Present	▪ None subscales ▪ Items represent different Qol domains	13	4 point (poor, fair, good, excellent)	Total	[9]	[21, 57, 64-81]
Quality of Life in Alzheimer's Disease Nursing home version, QoL-AD NH (proxy)	Mild – severe	Present	▪ None subscales ▪ Items represent different Qol domains	15	4 point (poor, fair, good, excellent)	Total	[35]	[51-53]
Quality of Life Assessment Schedule, QOLAS (proxy)	Mild – moderate	Present	A: Physical B: Psychological C: Social/Family D: Usual activities E: Cognitive functioning	10	6 point (no problem – it could not be worse)	Total	[36]	---
<b>Proxy-rating only</b>								
Alzheimer Disease Related Quality of Life, ADRQL (Japanese version called AD-HRQL-J or later QLDJ)	Mild – severe	Previous two weeks	A: Social interaction B: Awareness of self C: Feelings and mood D: Enjoyment of activities E: Response to surroundings A (QLDJ): Interacting with surroundings B (QLDJ): Expressing self C (QLDJ): Experiencing minimum negative behavior	Depending on version between 40 to 48 (QLDJ = 20 Items from the AD-HRQL + 4 new Items)	Dichotomous (agree, disagree) 4 point (not at all applicable – very much applicable) (QLDJ)	Total and subscales	[37]	[35, 52, 53, 81, 83-86]

Measurement	Dementia severity	Time frame	Qol Domains/Subscales	No. of Items	Response options	Scoring procedure	Primary papers (development)	Secondary papers (modification & validation in other countries)
Community Dementia Quality of Life Profile, CDQLP	Mild – severe	Previous two weeks	A: Thinking and behavior B: Family and social life C: Physical activities D: Other aspects of daily living	33	4 point (not at all – always)	Total	[38]	---
Dementia Care Mapping, DCM	Mild – severe	Present observations	A: Well being B: Activity	26	6 point (extreme ill-being – extreme well being) for well being scale	Subscales	[39]	[53, 87-90]
Observing Quality of Life in Dementia, OQOLD (Observed Quality of Life in Dementia Advanced, OQOLDA)	Mild – moderate (severe)	Present observations	<ul style="list-style-type: none"> <li>▪ None subscales</li> <li>▪ Observation is based on verbal and nonverbal cues and indicators of engagement and affect (OQOLD)</li> <li>▪ subtle signs of positive and negative affect and engagement (OQOLDA)</li> </ul>	6	7 point (+3 = an extremely unpleasant experience – -3 = an extremely unpleasant experience)	Total	[40]	---
Psychological well-being in cognitively impaired persons, PWP-CIP	Mild – moderate	Previous 24h	A: Positive affect/interaction B: Negative affect/interaction	11	4 point (never – frequently)	Total	[41]	---
Quality of Life for Dementia, Qol-D	Moderate – severe	Present	A: Positive affect B: Negative affect and actions C: Ability of communication D: Restlessness E: Attachment with others F: Spontaneity and activity	31	4 point (none – frequent)	Subscales	[42]	---
Quality of Life in Late-stage Dementia Scale, QUALID	Moderate – severe	Previous week	<ul style="list-style-type: none"> <li>▪ None subscales</li> <li>▪ Items represent different Qol domains</li> </ul>	11	5 point (different frequency options)	Total	[43]	[91-93]

Measurement	Dementia severity	Time frame	Qol Domains/Subscales	No. of Items	Response options	Scoring procedure	Primary papers (development)	Secondary papers (modification & validation in other countries)
QUALIDEM	Moderate – severe (very severe)	Previous two weeks	A: Care relationship B: positive affect C: negative affect D: restless tense behavior F: social relations G: social isolation  Not for very severe dementia: E: positive self-image H: feeling at home I: having something to do	37 (18 for very severe dementia) + 3 additional items for further measurement development	4 point (never – often)	Subscales	[10]	[16, 94-96]
Vienna List	Severe	Present	A: Communication B: Negative affect C: Bodily contact D: Aggression E: Mobility	40	5 point (never – always)	Subscales	[44]	---

**Appendix B****Limits: English, German**

---

**The following search terms were entered as independent terms, text words, or MESH terms and later combined with the Boolean operator AND:**

---

- (1) ((Dementia [MeSH]) OR (Alzheimer\* [tiab]) OR (Demen\* [tiab]) OR (Cognition Disorders [MeSH]) OR (Mild Cognitive Impairment [MeSH]))

---

  - (2) ((Life quality [tiab]) OR (Well being [tiab]) OR (Wellbeing [tiab]) (HRQoL [tiab]) OR (QOL [tiab]) OR (Quality of Life [MeSH]))

---

  - (3) ((PROMS [tiab]) OR (Geriatric assessment [MeSH]) OR (Test [tiab]) OR (Scal\* [tiab]) OR (Questionnaires [MeSH]) OR (Measure\* [tiab]) OR (Instrument [tiab]) OR (Assessment [tiab]) OR (Outcome Assessment [Health Care] [MeSH]))
-

**Appendix C**

Guideline for the process of linguistic validation (Beaton et al. 2000)

<b>Steps</b>	<b>Description</b>	<b>Rating scheme</b>
Translation	Two (or more) translators should independently translate the source questionnaire. The translators should preferably be native speakers to the target language	+ Translation performed by at least two independent translators ? Doubtful translation procedure (e. g. no translator who was aware of the concept) - Translation performed by only one translator 0 No information about translation
Synthesis	The translators should synthesize the multiple translations to produce a consensus of the translations	+ Performed synthesis ? Doubtful design (e. g. without recording observer) 0 No synthesis about translation OR translation performed by only one translator
Back translation	Translators, blind to the original measurement, should translate the consensus translation back to the original language	+ Back translation performed by at least two independent translators ? Doubtful back translation procedure (e. g. translators were not mother tongues) - Back translation performed by only one translator 0 No information about back translation
Expert committee	The expert committee should consolidate all the versions of the measurement and develop what would be considered the prefinal version of the measurement for testing	+ Clearly reported the existence of an expert committee ? Doubtful design (e. g. committee without translators or methodologists) 0 No information about the expert committee
Pretesting	The prefinal measurement undergoes pilot testing with members of the target population	+ Performed pretesting ? Doubtful design (e. g. n < 30) 0 No information about pretesting

+ positive rating, - negative rating, 0 no information available, ? unclear.

**Appendix D**

Quality criteria for the reliability of dementia specific QoL measurements based on the quality appraisal tool for studies of diagnostic reliability (QAREL) by Lucas et al. (2010) and the Criteria for psychometric properties of Health Status Questionnaires by Terwee et al. (2007)

<b>Criterion</b>	<b>Definition</b>	<b>Quality rating</b>
Internal Consistency (IC)	Internal consistency is a measure of the homogeneity of a (sub-) scale. It indicates the extent to which items in a (sub-) scale are inter-correlated, thus measuring the same construct. Factor analysis should be applied to determine the dimensionality of the item – this is to determine whether or not they formed only one overall dimension or more than one.	+ Factor analyses performed on adequate sample size (7 * X items and $\geq 100$ ) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95 ? No factor analysis OR doubtful design or method - Cronbach's alpha(s) < 0.70 OR > 0.95 despite adequate design and method 0 No information found on internal consistency
Inter-Rater Reliability (IRR)	Agreement between independent raters/observers; assessed by ICCs	+ ICC or kappa $\geq 0.70$ ? Doubtful design or method (e. g. no reporting of sample or rater characteristics, no reporting about the blinding of raters) - ICC or kappa < 0.70, despite adequate design and method 0 No information found on reliability
Test-Retest Reliability (TRR)	Stability assessed by administering the measurement to respondents on two different time points and examining the correlation between test and retest scores	+ ICC or kappa $\geq 0.70$ ? Doubtful design or method (e. g. no reporting of sample or rater characteristics, time interval not mentioned) - ICC or kappa < 0.70, despite adequate design and method 0 No information found on reliability

+ positive rating, - negative rating, 0 no information available, ? unclear.

## Appendix E: Quality of the linguistic validation process of dementia-specific QoL measurements

Source/ Language	Objective	Quality of linguistic validation					Overall conclusion	Methodological comments
		Trans- lation	Synthesis	Back translation	Committee review	Pilot testing		
<b>Dementia Quality of Life (DQoL) – Original [US]</b>								
Smith et al 2005 [21] [US] -> [GB]	Psychometric evaluation of the DEMQOL and DEMQOL-proxy in a final field test (DQoL scores were compared to DEMQOL scores)	0	0	0	0	0	?	<ul style="list-style-type: none"> <li>No information about linguistic validation of the US-version for the application in the GB reported</li> </ul>
Karim et al., 2008 [46] [US] -> [GB]	Cultural adaption and psychometrical evaluation from the US-version to the GB-version	0	0	0	+	?	I	<ul style="list-style-type: none"> <li>No linguistic validation of the US-version for the application in the GB reported</li> <li>Sample for the pilot test was n &lt; 30</li> </ul>
Chiu et al., 2010 [47] [US] -> [TW]	Translation and psychometric evaluation of the Chinese DQoL version	-	0	-	?	?	I	<ul style="list-style-type: none"> <li>Translation and back-translation were performed by only one translator</li> <li>No information about committee review reported</li> <li>Sample for the pilot test was n &lt; 30</li> </ul>
Wolak et al. 2010 [48] [US] -> [FR]	Contribution to the knowledge about the psychometric properties of the French DQoL version	0	0	0	0	0	?	<ul style="list-style-type: none"> <li>An existing French DQoL version was used.</li> <li>No information about the quality of linguistic validation reported</li> </ul>
Lucas-Carrasco et al., 2011 [49] [US] -> [ES]	Cultural adaption and psychometrical evaluation of the Spanish DQoL version	+	?	-	0	0	I	<ul style="list-style-type: none"> <li>Only one back-translator</li> <li>No information about a recording observer of the synthesis step</li> <li>No information about committee review and pilot testing reported</li> </ul>
Moyle et al. 2012 [51] [US] -> [AU]	Comparative analysis of the DQoL and QoL-AD-self NH as QoL instruments in nursing homes	0	0	0	0	0	?	<ul style="list-style-type: none"> <li>No information about linguistic validation of the US-version for the application in AU reported</li> </ul>
Voigt-Radloff et al., 2012 [50] [NL] -> [DE]	DQoL translation from Dutch into German and evaluation of reliability and validity	+	+	+	+	?	G	<ul style="list-style-type: none"> <li>No information about the linguistic validation of the Dutch DQoL version reported</li> <li>Sample for the pilot test was n &lt; 30</li> </ul>
<b>Cornell-Brown Scale for Quality of Life in Dementia (CBS QoL) – Original [US]</b>								
Lucas-Carrasco et al., 2013 [58]	Adaption and validation of the CBS original version in Spanish culture	+	+	-	?	0	M	<ul style="list-style-type: none"> <li>Only one back-translator</li> <li>No information about the members and</li> </ul>

Source/ Language	Objective	Quality of linguistic validation					Overall conclusion	Methodological comments
		Trans- lation	Synthesis	Back translation	Committee review	Pilot testing		
[US] -> [ES]								the purpose of the expert committee ▪ No information about pilot testing
<b>DEMQOL (proxy) – Original [GB]</b>								
Berwig et al. 2009 [60] [GB] -> [DE]	Evaluation of the effect of anosognosia and cognitive deficits on the reliability and validity of QoI	0	0	0	0	0	?	▪ A linguistic validated version was used [62]. This conference abstract is not accessible so the quality of linguistic validation is not clear
Lucas-Carrasco et al. 2010 [61] [GB] -> [ES]	Translation and psychometric evaluation of DEMQOL and DEMQOL-proxy	-	0	-	?	0	I	▪ No information about the qualification of translators and back-translators, only one translator and back-translator ▪ No information about participants of committee review ▪ No information about synthesis and pilot testing
<b>Quality of Life in Alzheimer's Disease, QoI-AD (proxy) – Original [US]</b>								
Selai et al. 2001 [64] [US] -> [GB]	Cross cultural adaption of the Original US-version of the QoI-AD to GB culture	0	0	0	0	?	I	▪ No linguistic validation of the US-version for the application in the GB reported ▪ Sample for the pilot test was n < 30
Novelli et al. 2005 [65] [US] -> [BR]	Cross-cultural adaption and psychometric evaluation of the Portuguese-version (Brazilian) of the QoI-AD scale	?	+	?	?	+	M	▪ No information about the qualification of translators and back-translators ▪ Translators and back-translators were not involved in the expert committee
Fuh et al. 2006 [66] [US] -> [TW]	Examination of the psychometric properties of the QoI-AD in Taiwan	0	0	0	0	0	?	▪ No information about linguistic validation of the US-version for the application in Taiwan reported
Matsui et al. 2006 [67] [US] -> [JP]	Development of the Japanese version of the QoI-AD and investigation of its psychometric properties	?	0	+	0	0	I	▪ No information about the number of translators for translation and their qualification ▪ No information about synthesis, committee review and pilot testing reported
Lin Kiat Yap et al. 2008 [68] [US] -> [CN-]	Examination of the psychometric properties and clinical experience in using the Mandarin translation of the	+	+	+	?	?	M	▪ No reporting about the involvement of translators and back-translators in the expert committee

Source/ Language	Objective	Quality of linguistic validation					Overall conclusion	Methodological comments
		Trans- lation	Synthesis	Back translation	Committee review	Pilot testing		
Mandarin]	QoI-AD							<ul style="list-style-type: none"> <li>▪ Sample for the pilot test was n &lt; 30</li> </ul>
Chan et al. 2011 [69] [CN-Mandarin] -> [CN- Cantonese]	Investigation of the psychometric properties and effects of cognitive function as well as depressive mood on the QoI of Hong Kong Chinese Alzheimer's disease patients	?	0	0	?	?	<b>I</b>	<ul style="list-style-type: none"> <li>▪ Translation was performed by only one translator</li> <li>▪ No information about synthesis and back-translation</li> <li>▪ No information about the participation of the translator and back-translators in the committee review reported</li> <li>▪ Pilot testing is reported without further description</li> </ul>
Wolak et al. 2009 [70] [US] -> [FR]	Evaluation of psychometric properties and the cross-cultural adaption into French of the quality of life in Alzheimer's disease scale	+	+	-	+	+	<b>G</b>	<ul style="list-style-type: none"> <li>▪ Only one back-translator reported</li> </ul>
Leon-Salas et al. 2011 [80] [US] -> [ES]	Evaluation of the psychometric properties of the Spanish version of the QoI-AD in institutionalized patients and families	0	0	0	0	0	?	<ul style="list-style-type: none"> <li>▪ No information which QoI-AD version were used</li> </ul>
Gomez-Gallego et al. 2014 [71] [US] -> [ES]	Investigation of the factorial structure of the QoI-AD in patients with Alzheimer's disease	+	?	?	?	0	<b>I</b>	<ul style="list-style-type: none"> <li>▪ No information about a recording observer of the synthesis step</li> <li>▪ No information about the number of translators for back-translation and their qualification</li> <li>▪ No information about the involvement of translators in committee review</li> <li>▪ No information about pilot testing reported</li> </ul>
Akpınar et al. 2012 [72] [US] -> [TR]	Investigation of the validity and reliability of the Turkish version of the QoI-AD	+	+	?	+	?	<b>M</b>	<ul style="list-style-type: none"> <li>▪ Only one back-translator</li> <li>▪ Sample for the pilot test was n &lt; 30</li> </ul>
Barrios et al. 2013 [73] [US] -> [PT]	Translation, adaption and validation of the QoI-AD to European Portuguese	+	+	+	?	?	<b>M</b>	<ul style="list-style-type: none"> <li>▪ No information about the participation of translators and back translators in committee review</li> <li>▪ Sample for the pilot test was n &lt; 30</li> </ul>

Source/ Language	Objective	Quality of linguistic validation					Overall conclusion	Methodological comments
		Trans- lation	Synthesis	Back translation	Committee review	Pilot testing		
Gräske et al. 2014 [81] [?] -> [DE]	Evaluation of psychometric properties of 3 Qol instruments in German shared housing arrangements	0	0	0	0	0	?	▪ No information which Qol-AD-proxy version were used
<b>Quality of Life in Alzheimer's Disease, QOL-AD (proxy) Nursing home version – Original [US]</b>								
Moyle et al. 2012 [51] [US] -> [AU]	Comparative analysis of the DQol and Qol-AD-self NH as Qol instruments in nursing homes	0	0	0	0	0	?	▪ No information about linguistic validation of the US-version for the application in AU reported
<b>Proxy-rating only</b>								
<b>Alzheimer Disease Related Quality of Life, ADRQL (Japanese version called AD-HRQL-J or later QLDJ) – Original [US]</b>								
Yamamoto- Mitani et al., 2000 [84] [US] -> [JP]	Examination of the psychometric properties of the Japanese translation of the ADRQL	?	0	?	0	0	I	▪ No information about the number of translators for translation and back- translation and their qualification ▪ No information about synthesis, committee review and pilot testing reported
Menzi-Kuhn (2006) [83] [US] -> [CH]	Investigation of the Quality of life of people with dementia living in nursing homes in Switzerland	-	0	-	?	?	I	▪ Translation and back-translation was performed by only one translator ▪ No information about the participation of translators and back translators in committee review ▪ Sample for the pilot test was n < 30
Gräske et al. 2013 [81] [US] -> [DE]	Evaluation of psychometric properties of 3 Qol instruments in German shared housing arrangements	0	0	0	0	0	?	▪ No information which ADRQL version were used
<b>Dementia Care Mapping, DCM – Original [GB]</b>								
Rüsing 2004 [89] [GB] -> [DE]	Evaluation of the Inter-rater reliability of Dementia Care Mapping	-	0	-	0	0	I	▪ Translation was performed by only one translator and a back-translation was not conducted ▪ No information about expert committee and pilot testing reported
Sloane et al. 2005 [53] [GB] -> [US]	Investigation of psychometric properties of Qol measures in long-term care facilities	0	0	0	0	0	?	▪ No information about linguistic validation of the GB-version for the application in US reported

Source/ Language	Objective	Quality of linguistic validation					Overall conclusion	Methodological comments
		Trans- lation	Synthesis	Back translation	Committee review	Pilot testing		
<b>Quality of Life in Late-stage Dementia Scale, QUALID – Original [US]</b>								
Falk et al. 2007 [91] [US] -> [SE]	Evaluation of the validity, reliability and responsiveness of a Swedish version of the QUALID	?	0	?	0	0	I	<ul style="list-style-type: none"> <li>▪ No information about the number of translators for translation and back-translation and their qualification</li> <li>▪ No information about synthesis, expert committee and pilot testing reported</li> </ul>
Garre-Olmo et al. 2010 [92] [US] -> [ES]	Development and evaluation the psychometric properties of the Spanish-version of the QUALID	+	+	+	?	0	M	<ul style="list-style-type: none"> <li>▪ No information about the participation of translators and back translators in committee review</li> <li>▪ No information about pilot testing reported</li> </ul>
Mjorud et al. 2014 [93] [US] -> [NO]	Investigation of the validity and internal consistency of the Norwegian QUALID-version	0	0	0	0	0	?	<ul style="list-style-type: none"> <li>▪ No information about the process of cross-cultural adaption reported, only citation of a not accessible conference abstract [99]</li> </ul>
<b>QUALIDEM – Original [NL]</b>								
Dichter et al. 2011 [94] [NL] -> [DE]	Investigation of the validity and reliability of the German version of the QUALIDEM	?	0	?	0	0	I	<ul style="list-style-type: none"> <li>▪ No information about the number of translators for translation and back-translation and their qualification</li> <li>▪ No information about synthesis, expert committee and pilot testing reported</li> </ul>

Overall conclusion: G = Good; M = Moderate; I = Insufficient; ? = No information reported or cited source not available.

## Appendix F: Quality and results of the reliability evaluation of dementia-specific Qol measurements

Source	Qol-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
<b>Self-rated only</b>							
<b>Bath Assessment of Subjective Quality of Life in Dementia (BASQID)</b>							
Trigg et al. 2007 [30]	BASQID (GB) ▪ 24 items	item reduction and evaluation of reliability during field testing (first field test)	<ul style="list-style-type: none"> <li>Consecutive sample of n = 55 people with dementia recruited from a memory clinic</li> <li>Inclusion (participants): Dementia diagnosis based on DSM IV criteria, MMSE score <math>\geq</math> 12</li> <li>Exclusion (participants): English not first language</li> </ul>	IC	0.91	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>No analysis of unidimensionality</li> </ul>
			<ul style="list-style-type: none"> <li>Convenience sample of n = 29 people of the original sample, all other sample information was equal to the IC sample</li> </ul>	TRR	0.82 (0.66-0.91)	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>n = 1 data sets were excluded due to missing values</li> <li>Second observation after 2 weeks</li> </ul>
Trigg et al. 2007 [45]	BASQID (GB) ▪ 14 +3 items  A: Life satisfaction (8) B: Feelings of positive Qol (6) + 3 additional items  Original version (GB)	Development of a new Qol measure and evaluation of reliability and validity of the new measure (second field test)	<ul style="list-style-type: none"> <li>Consecutive sample of n = 143 people with dementia recruited from a memory clinic</li> <li>Inclusion (participants): dementia diagnosis based on DSM IV criteria, MMSE <math>\geq</math> 12 or a clinical judgment that clinical participants were able to complete the BASQID</li> <li>Exclusion (participants): English not first language, participation on the first BASQID field test</li> </ul>	IC	A: 0.84 B: 0.83	+	<ul style="list-style-type: none"> <li>Subscales are based on a conducted factor analysis</li> </ul>
			<ul style="list-style-type: none"> <li>Consecutive subsample of n = 29 people with dementia for TRR from the sample of the first field test</li> <li>All sample information was equal to the first field test</li> </ul>	TRR	A: 0.79 (0.59-0.89) B: 0.85 (0.70-0.93)	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>Second observation after 2 weeks</li> <li>No information about blinding reported</li> </ul>
<b>Dementia Quality of life (DQoI)</b>							
Brod et al. 1999 [11]	DQoI (US) ▪ 29 +1 items  A: Self-esteem (4) B: Positive affect/humour (6) C: Negative affect (11) D: Feelings of belonging (3) E: Sense of aesthetics (5) + 1 additional item rating overall Qol	Development, pilot testing and field testing of the DQoI	<ul style="list-style-type: none"> <li>Convenience sample of n = 95 people with dementia living at home</li> <li>Inclusion (participants): dementia diagnosis, English speaking, having a spouse or related caregiver who either lived with them or visited at least three times per week</li> <li>Exclusion (Participants): Unable to complete filter questions (n = 4)</li> </ul>	IC	A: 0.80 B: 0.83 C: 0.89 D: 0.67 E: 0.77	?	<ul style="list-style-type: none"> <li>Subscales were not based on a factor analysis</li> </ul>
			<ul style="list-style-type: none"> <li>Convenience sample of n = 18 people of the original sample, all other sample information was equal to the IC sample</li> </ul>	TRR	A: 0.68 <sup>a</sup> B: 0.90 C: 0.64 D: 0.74 E: 0.72	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>No information about blinding reported</li> <li>Only total ICC without CI were reported</li> <li>No appropriate statistical measure of agreement used</li> <li>Second observation after 2 weeks</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Original version (US)							
Edelman et al. 2004 [52]	DQoI (US) ▪ 29 +1 items ▪ Original subscales	Comparison of the properties of different dementia-specific QoI measurements	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 36 people with dementia recruited from 3 adult day care centres</li> <li>▪ Inclusion (participants): MMSE <math>\geq</math> 10</li> <li>▪ Exclusion (participants): Not reported</li> </ul>	IC	0.84	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Results for internal consistency were not based on the results of a factor conducted analysis</li> <li>▪ Doubtful calculation of Cronbach's alpha according to the multidimensionality of the DQoI</li> </ul>
Sloane et al. 2005 [53]	DQoI (US) ▪ 29 +1 items ▪ Original subscales	Investigation of psychometric properties of QoI measures in long-term care facilities	<ul style="list-style-type: none"> <li>▪ Purposive sample of n = 100 people with dementia recruited from residential care facilities, assisted living facilities and nursing homes</li> <li>▪ Inclusion (Participants): Age <math>\geq</math> 65 years, diagnosis of dementia</li> <li>▪ Exclusion (participants): Primary diagnosis of Huntington's disease, alcohol related dementia, schizophrenia, manic-depressive disorders, mental retardation, unable to complete 3 DQoI screening items</li> </ul>	IC	A: 0.73 B: 0.84 C: 0.88 D: 0.59 E: 0.83	?	<ul style="list-style-type: none"> <li>▪ Small sample (n = 97 to 100 in subscale)</li> <li>▪ Results for internal consistency were not based on the results of a conducted factor analysis</li> <li>▪ No characteristics for the subsample of self-rating participants reported</li> </ul>
Edelman et al. 2005 [35]	DQoI (US) ▪ 29 +1 items ▪ Original subscales	Investigation of the distribution and factor structure of the items and psychometric properties of the dementia-specific QoI instrument	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 58 people with dementia recruited from 4 dementia specific nursing homes and 3 assisted living facilities</li> <li>▪ Inclusion (participants): MMSE <math>\geq</math> 10</li> <li>▪ Exclusion (participants): Not reported</li> </ul>	IC	A: 0.70 B: 0.78 C: 0.95 D: 0.61 E: 0.79	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Results for internal consistency were not based on the results of a conducted factor analysis</li> </ul>
Ready et al. 2006 [54]	DQoI (US) ▪ 29 +1 items ▪ Original subscales	Determine if patient insight and cognitive impairment are factors that can guide decisions regarding the reliability and validity of self-report data from cognitively impaired persons	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 68 participants, the majority was recruited from a memory disorder clinic</li> <li>▪ Inclusion (participants): Age <math>\geq</math> 60 years, 6 years of education, diagnosis of mild Alzheimer's disease based on the NINCDS-ADRDA criteria or mild cognitive impairment, dyad of participant and caregiver where the caregiver is also willing to participate</li> <li>▪ Exclusion (participants): Other neurologic disorders besides cognitive impairment, major DSM-IV diagnosis within the past two years, current alcohol or substance abuse, or living in an institutional setting</li> </ul>	IC	Insight: low / high A: 0.65 / 0.75 B: 0.61 / 0.85 C: 0.76 / 0.88 D: 0.63 / 0.58 E: 0.55 / 0.70	?	<ul style="list-style-type: none"> <li>▪ Small sample size (n = 35 low insight, n = 33 high insight)</li> <li>▪ Cronbach's alpha values were calculated for two subsamples of people with low insight or high insight</li> <li>▪ Calculation of internal consistency was not based on the results of a conducted factor analysis</li> </ul>
Adler et al. 2010 [55]	DQoI (US) ▪ 29 +1 items ▪ Original subscales	Evaluation of the psychometric properties of the DQoI in nursing	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 486 people with dementia recruited from nursing homes</li> <li>▪ Inclusion (participants): Age <math>\geq</math> 65 years, MMSE score <math>\geq</math> 11,</li> </ul>	IC	A: 0.62 B: 0.72 C: 0.84	+	<ul style="list-style-type: none"> <li>▪ Subscales are based on the results of a confirmatory factor analysis</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
		home residents	life expectancy > 6 months, not receiving skilled rehabilitation services <ul style="list-style-type: none"> <li>Exclusion (participants): Not reported</li> </ul>		D: 0.53 E: 0.65		
			<ul style="list-style-type: none"> <li>n = 231 from the original sample</li> <li>Selection criteria of the TRR sample were the assignment to the control group. All sample information was equal to the IC sample</li> </ul>	TRR	A: 0.31 (0.74) <sup>b</sup> B: 2.00 (0.13) C: 0.56 (0.57) D: 1.50 (0.21) E: 0.88 (0.42)	?	<ul style="list-style-type: none"> <li>No sample size calculation</li> <li>Only comparison of mean values at baseline, after 4 months and 12 months (no significant difference identified)</li> <li>No use of appropriate statistical measures like ICC</li> <li>No information about blinding reported</li> </ul>
Moyle et al. 2012 [51]	DQoI (US) <ul style="list-style-type: none"> <li>29 +1 items</li> <li>Original subscales</li> </ul>	Comparative analysis of the DQoI and QoI-AD-self NH as QoI instruments in nursing homes	<ul style="list-style-type: none"> <li>Convenience sample of n = 42 people with dementia recruited from 4 long-term care facilities</li> <li>Inclusion (participants): Confirmed diagnosis of dementia or probable dementia based on a MMSE score between 10 to 24 or features consistent with the DSM-IV criteria, ability to complete surveys with the aid of an assistant, one family member or friend who was accessible and in contact with the participant</li> <li>Exclusion (participants): Not reported</li> </ul>	IC	A: 0.64 B: 0.58 C: 0.79 D: 0.48 E: 0.70	?	<ul style="list-style-type: none"> <li>The original US-version of the DQoI was used in an AU sample without cross-cultural adaption</li> <li>Small sample</li> <li>The calculation of Internal consistency was not based on the results of a factor analysis</li> </ul>
Smith et al. 2005 [21]	DQoI (US) <ul style="list-style-type: none"> <li>29 +1 items</li> <li>Original subscales</li> </ul>	Psychometric evaluation of the DEMQoI and DEMQoI-proxy in a final field test (DQoI scores were compared to DEMQoI scores)	<ul style="list-style-type: none"> <li>Random sample of n = 30 people with dementia based on a convenience sample of n = 75 people with dementia recruited from community mental health teams and Healthcare National Health Service Trust</li> <li>Inclusion (participants): dementia diagnosis based on ICD-10 criteria, MMSE <math>\geq</math> 12</li> <li>Exclusion (participants): Pass more than 1 Screening-Question of The DQoI</li> </ul>	IC	A: 0.87 B: 0.78 C: 0.83 D: 0.55 E: 0.37	-	<ul style="list-style-type: none"> <li>The original US-version of the DQoI was used in a UK sample without cross-cultural adaption</li> <li>Small sample size</li> <li>Subscales were not based on a factor analysis</li> <li>Depending of the subscale n = 12 to 15 data sets were excluded due to missing values</li> </ul>
Karim et al. 2008 [46]	DQoI (GB) <ul style="list-style-type: none"> <li>29 +1 items</li> <li>Original subscales</li> </ul>	Translation and psychometrical evaluation of the DQoI GB-version	<ul style="list-style-type: none"> <li>Convenience sample of n = 36 people with ICD-10 diagnosis of dementia recruited from a hospital-based psychiatric service</li> <li>Inclusion (participants): mild to moderate dementia</li> <li>Exclusion (participants): Not reported</li> </ul>	IC	A: 0.73-0.74 <sup>c</sup> B: 0.74-0.82 C: 0.76-0.80 D: 0.56-0.71 E: 0.78-0.79	?	<ul style="list-style-type: none"> <li>Small sample</li> <li>Cronbach's alpha values were only reported for each item</li> <li>The calculation of Internal consistency was not based on the results of a factor analysis</li> </ul>
			<ul style="list-style-type: none"> <li>Sample information was equal to the IC sample</li> </ul>	TRR	A: 0.76 <sup>d</sup> B: 0.71 C: 0.75 D: 0.71 E: 0.71	?	<ul style="list-style-type: none"> <li>Second observation after 2 weeks</li> <li>No sample size calculation</li> <li>Only total Spearman correlation coefficients without CI were reported</li> <li>No information about blinding reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Chiu et al. 2008 [56]	DQoI (TW) ▪ 29 +1 items ▪ Original subscales	Validation of the Taiwan DQoI in early to moderate Alzheimer's disease, comparison of differences in DQoI across three dementia severity groups and exploration of related factors of DQoI values	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 36 people with Alzheimer's disease based on DSM-IV criteria, n = 35 people with questionable dementia and n = 27 age- and education-matched controls without cognitive symptoms from northern Taiwan</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li> </ul>	IC	0.82-0.92 <sup>c</sup> (range)	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ Calculation of internal consistency was not based on the results of the conducted factor analysis</li> <li>▪ Cronbach's alpha values are not reported for each subscale as Range</li> <li>▪ Sample also consists of people without dementia</li> <li>▪ No information about the translation and cross-cultural adaption process of the used Chinese DQoI version</li> </ul>
Chiu et al. 2010 [47]	DQoI (TW) ▪ 29 +1 items ▪ Original subscales	Test of psychometric properties of the Chinese DQoI	<ul style="list-style-type: none"> <li>▪ Consecutive sample of n = 110 participants (N = 44 people with Alzheimer's disease and n = 39 people with questionable dementia and n = 27 people without dementia recruited in 2 memory disorder clinics</li> <li>▪ Inclusion (participants): Mandarin Chinese, Taiwanese or Hakka dialect speakers with a main primary caregiver, DSM-IV and NINCDS-ADRDA diagnostic criteria</li> <li>▪ Exclusion (participants): Acute illnesses, impaired sensory problems (hearing loss and severe visual problems), chronic alcohol abuse, use of drugs which possibly affect central nervous system functions, CDR &gt; 2, MMSE &lt; 10, incomplete demographic data</li> </ul>	IC	A: 0.84 B: 0.89 C: 0.94 D: 0.53 E: 0.86	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ Evaluation of internal consistency was not based on the results of a factor analysis</li> <li>▪ Sample also consists of people without dementia</li> </ul>
Wolak et al. 2010 [48]	DQoI (FR) ▪ 29 +1 items ▪ Original subscales	Contribution to the knowledge about the psychometric properties of the French DQoI version	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 109 people with dementia recruited in a geriatric department or a memory clinic in France or French-speaking Switzerland</li> <li>▪ Inclusion (participants): Age <math>\geq</math> 65 years, DSM-IV diagnostic criteria, native French speakers with a main caregiver, MMSE &gt; 10</li> <li>▪ Exclusion (participants): More than one incorrect answer in a screening test</li> </ul>	IC	A: 0.72 B: 0.85 C: 0.87 D: 0.70 E: 0.75	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ No information about cross-cultural adaptation reported</li> <li>▪ Evaluation of internal consistency was not based on the results of a factor analysis</li> </ul>
			<ul style="list-style-type: none"> <li>▪ Sample information was equal to the IC sample</li> </ul>	TRR	A: 0.97 B: 0.99 C: 1.00 D: 0.99 E: 0.96	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation</li> <li>▪ Only total ICC without CI was reported</li> <li>▪ No information about blinding reported.</li> <li>▪ Second observation after 2 weeks</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Wolak-Thierry et al. 2014 [57]	DQoI (FR) ▪ 29 +1 items ▪ Original subscales	Comparison of psychometric properties of the DQoI and QOL-AD to identify which instrument is most suitable for the use for consultations in geriatric medicine	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 121 people with Alzheimer's disease recruited in a geriatric department or a memory clinic</li> <li>▪ Inclusion (participants): Dementia diagnosis based on DSM-IV and NINCDS-ADRDA criteria, age <math>\geq</math> 65 years, MMSE <math>\geq</math> 10</li> <li>▪ Exclusion (participants): Not reported</li> </ul>	IC	A: 0.72 B: 0.83 C: 0.84 D: 0.71 E: 0.74	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ Evaluation of internal consistency was not based on the results of a factor analysis</li> </ul>
	DQoI (FR) ▪ 29 +1 items ▪ Original subscales		<ul style="list-style-type: none"> <li>▪ Sample information was equal to the IC sample</li> </ul>	TRR	A: 0.71 B: 0.80 C: 0.74 D: 0.70 E: 0.69	?	<ul style="list-style-type: none"> <li>▪ Second observation after 14 days</li> <li>▪ Only total ICC without CI was reported</li> <li>▪ No information about blinding reported</li> <li>▪ No sample size calculation</li> <li>▪ No information about excluded participants reported</li> </ul>
Lucas-Carrasco et al. 2011 [49]	DQoI (ES) ▪ 29 +1 items ▪ Original subscales	Cultural adaption and psychometrical evaluation of the Spanish DQoI version	<ul style="list-style-type: none"> <li>▪ Convenience sample n = 101 recruited from diagnostic dementia units, day centers and day hospitals</li> <li>▪ Inclusion(participants): Dementia diagnosis based on DSM-IV criteria, a MMSE <math>\geq</math> 10, were living at home, had a known caregiver</li> <li>▪ Exclusion (participants): unable to complete the 3 DQoI screening items</li> </ul>	IC	A: 0.62 B: 0.76 C: 0.83 D: 0.62 E: 0.79	?	<ul style="list-style-type: none"> <li>▪ Small sample (calculated 4 subject per item)</li> <li>▪ Evaluation of internal consistency was not based on the results of a factor analysis</li> </ul>
			<ul style="list-style-type: none"> <li>▪ TRR: n = 28 people with dementia</li> <li>▪ Selection criteria for TRR sample were not reported</li> </ul>	TRR	A: 0.58 B: 0.90 C: 0.81 D: 0.67 E: 0.76	-	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ No information about blinding reported</li> <li>▪ Only total ICC without CI was reported</li> <li>▪ Second observation after 2 weeks <math>\pm</math> 3 days</li> </ul>
Voigt-Radloff et al. 2012 [50]	DQoI (DE) ▪ 29 +1 items ▪ Original subscales	DQoI translation into German and evaluation of reliability and validity	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 286 people with dementia living in the community</li> <li>▪ Inclusion (participants): Alzheimer's disease, vascular dementia based on DSM-IV criteria, MMSE score: 12 to 26, living in the community, primary carer available at least twice a week</li> <li>▪ Exclusion (participants): Major depression or behavioral disturbances</li> </ul>	IC	A: 0.68 B: 0.84 C: 0.84 D: 0.62 E: 0.67	?	<ul style="list-style-type: none"> <li>▪ Sample consists of two pooled samples of two randomized controlled trials</li> <li>▪ Evaluation of internal consistency was not based on the results of a factor analysis</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
<b>Mild Cognitive Impairment Questionnaire (MCQ)</b>							
Dean et al. 2014 [31]	MCQ (GB) ▪ 13 items  A: Emotional effects (6) B: Practical concerns (7)  Original version (GB)	Development and psychometric evaluation of the MCQ	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 146 people with mild cognitive impairment recruited from 11 research databases and living at home</li> <li>▪ Inclusion (participants): Age <math>\geq</math> 50 years, diagnosis of mild cognitive impairment confirmed by a clinician at a secondary care memory service in the 12 months prior to the QoI rating, able to read and write in English language</li> <li>▪ Exclusion (participants): Not reported</li> </ul>	IC	A: 0.91 B: 0.85	+	<ul style="list-style-type: none"> <li>▪ The evaluation of internal consistency was based on the results of a conducted explanatory factor analysis</li> </ul>
<b>Self-rating and proxy rating</b>							
<b>Activity and Affect Indicator Quality of life, AAIQOL (proxy)</b>							
Albert et al. 1996 [32]	AAIQOL (US) ▪ 21 items  A: Affect (6) B: Activity (15)  Original version (GB)	Examination of behaviors indicative for people with Alzheimer's disease and correlates of this behavior (quality of life)	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 10 people with dementia recruited from the community and nursing homes</li> <li>▪ Inclusion (participants): Dementia diagnosis based on DSM-III criteria and NINDS-ADRDA criteria for probable Alzheimer's disease, MMSE score <math>\geq</math> 30</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Data were collected by trained interviewers during a telephone interview</li> </ul>	TRR	A: 0.53 <sup>e</sup> to 0.92 (proxy) B: Frequency of activity > 0.40 for 13 of 15 Items, Enjoyment of activity > 0.70 for 14 of 15 Items (proxy)	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ No characteristics for the sample of people with dementia and proxy-raters reported</li> <li>▪ Only total Kappa values without CI were reported, Kappa values were calculated for each item</li> <li>▪ No information about blinding reported</li> <li>▪ Second observation after one week</li> </ul>
Sloane et al. 2005 [53]	AAIQOL(US) ▪ 21 items  A: Positive Affect (3) B: Negative Affect (3) C: Activity (15)	Investigation of psychometric properties of QoI measures in long-term care facilities	<ul style="list-style-type: none"> <li>▪ Purposive sample of n = 110 people with dementia recruited from residential care facilities, assisted living facilities and nursing homes (for the QoI-D-self ratings)</li> <li>▪ Purposive sample of n = 403 people with dementia recruited from residential care facilities, assisted living facilities and nursing homes (for the QoI-D-proxy ratings)</li> <li>▪ Inclusion (participants): Age <math>\geq</math> 65 years, diagnosis of dementia, MMSE <math>\geq</math> 10</li> <li>▪ Exclusion (participants): Primary diagnosis of Huntington's disease, alcohol related dementia, schizophrenia, manic-depressive disorders, mental retardation</li> <li>▪ Inclusion (proxies): Registered nurses or nursing assistant who know the resident best</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	C: 0.78  A: 0.66 (proxy) B: 0.66 (proxy) C: 0.78 (proxy)	?	<ul style="list-style-type: none"> <li>▪ The evaluation of internal consistency was not based on the results of a factor analysis</li> <li>▪ No characteristics for the subsample of self-rating participants and proxy-raters reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
			<ul style="list-style-type: none"> <li>Convenience sample of n = 23 people of the IC sample, all other sample information was equal to the IC sample</li> </ul>	IRR	A: 0.99 (proxy) B: 0.96 (proxy) C: 0.95 (proxy)	?	<ul style="list-style-type: none"> <li>No sample size calculation conducted (n = 18-23 per subscale)</li> <li>No selection criteria for the IRR sample reported</li> <li>No characteristics for the sample of people with dementia and proxy-raters reported</li> <li>Only total ICC values without CI were reported</li> <li>No information about blinding reported</li> </ul>
<b>Cornell Brown Scale for Quality of life in dementia, CBS QoI (proxy)</b>							
Ready et al. 2002 [33]	CBS QoI (US) 19 Items No subscales  Original version (US)	Development and psychometric evaluation of the CBS QoI scale	<ul style="list-style-type: none"> <li>Convenience sample of n = 50 people with dementia based on DSM-IV criteria or mild cognitive impairment recruited from a hospital-based outpatient dementia clinic</li> <li>Inclusion (participants): Not reported</li> <li>Exclusion (participants): Lack of ability to communicate basic needs, yes or no questions, unavailable reliable caregiver for joint interview</li> <li>Scale scores are based on clinical ratings, requiring joint interviews between clinicians, patients and caregivers</li> <li>Inclusion (proxies): Daily contact with people with dementia or mild cognitive impairment</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: 3 clinicians were included: 1 of the CBS QoI developers, 1 geriatric psychiatry fellow and 1 senior neurology resident the latter two were trained in the use of the CBS by one of the CBS QoI developers</li> </ul>	IC	0.81	?	<ul style="list-style-type: none"> <li>Small sample</li> <li>The evaluation of internal consistency was based on the results of a conducted explanatory factor analysis</li> <li>No information about caregivers and characteristics of proxy-raters reported</li> </ul>
			<ul style="list-style-type: none"> <li>Convenience sample n = 25 of the IC-sample, all sample information was equal to the IC-sample</li> </ul>	IRR	0.90 (proxy)	?	<ul style="list-style-type: none"> <li>Small sample</li> <li>No information about proxy characteristics reported</li> <li>Only total ICC without CI was reported</li> <li>No information about blinding reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Ready et al. 2008 [59]	CBS QoI (US) ▪ 19 items ▪ No subscales	Determination whether aggregating participant and informant quality-of-life reports on the CBS QoI can provide a broader perspective on the QoI relative to participant or caregiver/informant reports separately	<ul style="list-style-type: none"> <li>▪ Convenience sample n = 63 people with dementia or mild cognitive impairment and their caregiver respectively knowledgeable informant's (mild cognitive impairment) mainly recruited from a hospital-based memory disorder clinic</li> <li>▪ Inclusion (participants): Age <math>\geq</math> 60 years, 6 years of education, diagnosis of mild Alzheimer's disease based on NINCDS-ADRDA criteria or mild cognitive impairment, willingness to participate</li> <li>▪ Exclusion (participants): Other neurologic disorders besides cognitive impairment, major DSM-IV diagnosis within the past two years, current alcohol or substance abuse, or living in an institutional setting</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: QoI rating was guided by a clinician (one of the CBS QoI developers)</li> </ul>	IC	0.83  0.84 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ Evaluation of internal consistency was not based on a factor analysis</li> </ul>
Lucas-Carrasco et al. 2013 [58]	CBS QoI (ES) ▪ 19 items ▪ No subscales	Adaption and validation of the CBS QoI in Spanish culture	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 100 people with dementia recruited from 2 diagnostic dementia units and three day hospitals</li> <li>▪ Inclusion (participants): Dementia diagnosis based on DSM IV criteria, MMSE score <math>\geq</math> 10, living at home, people must have a known caregiver with daily contact and be able to answer 2 of 3 screening questions included in the DQoI</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: QoI ratings were guided by trained psychologists</li> </ul>	IC	0.87	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ Evaluation of internal consistency was not based on the results of the conducted factor analysis</li> <li>▪ No information about proxy characteristics reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
<b>DEMQOL (proxy)</b>							
Smith et al. 2005 [21]	DEMQOL (GB) ▪ 28 items  A: Daily activities (9) B: Memory (6) C: Negative emotion (8) D: Positive emotion (5) + 1 additional item	Development and psychometric evaluation (item reduction) of the DEMQOL and DEMQOL-proxy in a preliminary field test	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 130 people with dementia and their family carers n = 126 recruited from clinical psychiatry contacts in the UK</li> <li>▪ Inclusion (participants: Dementia diagnosis based on ICD-10 criteria</li> <li>▪ Exclusion (participants): Insufficient spoken English</li> <li>▪ Inclusion (proxies): Family carers of participants</li> <li>▪ Exclusion (proxies): Insufficient spoken English</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	A: 0.84 B: 0.89 C: 0.84 D: 0.85  A: 0.90 (proxy) B: 0.85 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ Evaluation of internal consistency was based on the results of a conducted factor analysis</li> </ul>
	DEMQOL (proxy), (GB) ▪ 31 items  A: Functioning (18) B: Emotion (11) + 3 additional items	Psychometric evaluation of the DEMQOL and DEMQOL-proxy in a final field test	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 75 people with dementia and their family carers n = 98 recruited from community mental health teams and Healthcare National Health Service Trust</li> <li>▪ Inclusion (participants): Dementia diagnosis based on ICD-10 criteria</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Family carers of the people with dementia</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	0.87  0.89 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ Evaluation of internal consistency was not based on the results of the conducted factor analysis</li> <li>▪ Doubtful calculation of Cronbach's alpha according to the multidimensionality of the DEMQOL and DEMQOL-proxy</li> </ul>
	Original version (GB)		<ul style="list-style-type: none"> <li>▪ Random sample of n = 17 people with dementia and n = 23 family carers from the final field test sample plus n = 7 additional recruited people with dementia and their family carers</li> <li>▪ All sample information was equal to the IC-sample</li> </ul>	TRR	0.84  0.75 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ No information about blinding reported</li> <li>▪ Only total ICC without CI were reported</li> <li>▪ Second observation after 14 days</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Berwig et al. 2009 [60]	DEMQOL (DE) <ul style="list-style-type: none"> <li>▪ 28 items</li> <li>A: Activities of daily living</li> <li>B: Memory</li> <li>C: Emotions</li> </ul> DEMQOL (proxy), (DE) <ul style="list-style-type: none"> <li>▪ 31 items</li> <li>A: Activities of daily living</li> <li>B: Memory</li> <li>C: Emotions</li> </ul>	Evaluation of the effect of anosognosia and cognitive deficits on the reliability and validity of QoI	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 27 people with Alzheimer's disease or mild cognitive impairment recruited from a memory clinic</li> <li>▪ Participants with high insight n = 14, participants with low insight n = 13</li> <li>▪ Inclusion (participants): Availability of a healthy caregiver for each study participant, age <math>\geq</math> 65 years</li> <li>▪ Exclusion (participants): Another neurological illness, current alcohol or drug abuse, further psychiatric diagnoses based on ICD-10 criteria</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	Insight: low / high A: 0.58 / 0.78 B: 0.67 / 0.83 C: 0.79 / 0.88 A: 0.83 / 0.86 (proxy) B: 0.57 / 0.84 (proxy) C: 0.82 / 0.89 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ Evaluation of internal consistency was not based on the results of the conducted factor analysis</li> </ul>
Lucas-Carrasco et al. 2010 [61]	DEMQOL (ES) <ul style="list-style-type: none"> <li>▪ 28 items</li> </ul> Factor analysis: <ul style="list-style-type: none"> <li>A: Everyday Life (6)</li> <li>B: Memory (8)</li> <li>C: Feelings (14)</li> </ul> DEMQOL (proxy), (ES) <ul style="list-style-type: none"> <li>▪ 31 items</li> </ul> Factor analysis: <ul style="list-style-type: none"> <li>A: Everyday life (6)</li> <li>B: Memory (7)</li> <li>C: Positive feelings (10)</li> <li>D: Negative feelings (8)</li> </ul>	Translation and psychometric evaluation of DEMQOL and DEMQOL-proxy from the GB version in Spanish culture	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 102 people with dementia recruited from dementia diagnostic units, day centers and day hospitals</li> <li>▪ Inclusion (participants): Diagnosis of dementia based on the DSM.IIR/DSM-IV Axis I criteria, a MMSE score <math>\geq</math> 10, a known caregiver, living at home</li> <li>▪ Exclusion (participants): MMSE score &lt; 10</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	A: 0.81 B: 0.82 C: 0.84 A: 0.80 (proxy) B: 0.86 (proxy) C: 0.84 (proxy) D: 0.73 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample</li> <li>▪ Evaluation of internal consistency was based on the results of the conducted factor analysis</li> </ul>
			<ul style="list-style-type: none"> <li>▪ People with dementia n = 25 of the IC- sample, all sample information was equal to the IC-sample</li> </ul>	TRR	0.71	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation</li> <li>▪ No analysis of TRR for DEMQOL-proxy</li> <li>▪ Only total ICC without CI were reported</li> <li>▪ No information about blinding reported</li> <li>▪ Second observation after 2 weeks <math>\pm</math> 3 days</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
<b>Heidelberg instrument for the assessment of quality of life in dementia (H.I.L.DE.)</b>							
Becker et al. 2011 [63]	H.I.L.DE. (DE) ▪ 48 items  A: Medical care and pain experience (4) B: Physical environment (10) C: Activities (8) D: Social reference system (5) E: Emotional-being and life satisfaction (14)	Development and first psychometric evaluation of the H.I.L.DE.	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 27 people with dementia and N = 14 caregivers from nursing homes</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li>   <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Proxies were caregivers</li> </ul>	IRR	Interval scaled items (proxy): A: 0.68 <sup>a</sup> B: 0.72-1.00 (range) C: 0.74-0.96 (range) D: 0.47-0.94 (range) E: 0.73-0.94 (range)  Categorical items (proxy): A: 0.78-1.00 <sup>a</sup> (range) B: 0.75 E: 0.53-1.00 (range)	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation</li> <li>▪ No characteristics of people with dementia and proxy-rating caregivers reported</li> <li>▪ Only total Pearson coefficients or Kappa values without CI were reported</li> <li>▪ No information about blinding reported</li> </ul>
			<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 26 people with dementia and n = 14 caregivers from nursing homes</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li>   <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Proxies were caregivers</li> </ul>	TRR	Interval scaled items (proxy): A: 0.68 <sup>a</sup> B: 0.73-1.00 (range) C: 0.42-0.91 (range) D: 0.46-0.81 (range) E: 0.38-0.84 (range)  Categorical items (proxy): A: 86-95% (range) B: 86% E: 52-100% (range)	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation</li> <li>▪ No characteristics of people with dementia and proxy-rating caregivers reported</li> <li>▪ No sample size calculation.</li> <li>▪ Only total Pearson coefficients without CI were reported</li> <li>▪ No information about blinding reported</li> <li>▪ Second observation after 5 days, additional evaluation was made after 8 days. Results were comparable but not displayed here</li> <li>▪ In some cases second observation was done by different proxy-rater (n = 5)</li> </ul>
<b>Quality of Life- Alzheimer's disease (QoI-AD), self- and proxy-rated version</b>							
Logsdon et al. 1999 [9]	QoI-AD (US) ▪ 13 items ▪ No subscales  Original version (US)	Development and psychometric evaluation of the QoI-AD	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 77 people with dementia recruited from a patient registry</li> <li>▪ Inclusion (participants): Diagnoses of probable or possible Alzheimer's disease based on NINCDS-ADRDA criteria, an actively involved caregiver</li> <li>▪ Exclusion (participants): Not reported</li>   <li>▪ Inclusion (proxies): who lived with the participant or spent every day with them</li> <li>▪ Exclusion (proxies): Not reported</li> </ul>	IC	0.88  0.87 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
			<ul style="list-style-type: none"> <li>Proxy-training: Interviews for self- and proxy-ratings were guided by an experienced interviewer, participants and proxies received explicit instructions</li> </ul>				
			<ul style="list-style-type: none"> <li>Convenience sample of n= 30 people of the IC sample, all other Sample information was equal to the IC sample</li> </ul>	TRR	0.76 0.92 (proxy)	?	<ul style="list-style-type: none"> <li>No sample size calculation conducted</li> <li>No selection criteria for the TRR sample reported</li> <li>No characteristics for the TRR subsample of people with dementia and proxy-raters reported</li> <li>Only total ICC values without CI were reported</li> <li>No information about blinding reported</li> <li>Second observation after 1 week</li> </ul>
Logsdon et al. 2002 [74]	QoI-AD (US) <ul style="list-style-type: none"> <li>13 items</li> <li>No subscales</li> </ul>	Evaluation of theoretical, psychometric and practical considerations of the QoI-AD self and proxy version	<ul style="list-style-type: none"> <li>Convenience sample of n = 155 people with dementia living in the community</li> <li>Inclusion (participants): Alzheimer's disease based on NINCDS-ADRDA criteria, community dwelling, availability of a actively involved caregiver</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (proxies): who lived with the participant or spent every day with them</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Not reported</li> </ul>	IC	0.84 0.86 (proxy)	?	<ul style="list-style-type: none"> <li>Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>
Hoe et al. 2005 [75]	QoI-AD (US) <ul style="list-style-type: none"> <li>13 items</li> <li>No subscales</li> </ul>	Examination of the usefulness of the QoI scale in Alzheimer's disease (QoI-AD US-version) in people with severe dementia	<ul style="list-style-type: none"> <li>Representative sample of n = 41 people with dementia living at home and in care homes</li> <li>Inclusion (participants): Alzheimer's disease based on DSM-IV and NINCDS-ADRDA criteria, MMSE &lt; 12</li> <li>Exclusion (participants): Cancel of the survey of distress</li> <li>Inclusion (proxies): Statutory carers which spend <math>\geq 4</math> hours a week with the participants</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Not reported</li> </ul>	IC	0.78 (2*self and 1*proxy)	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> <li>QoI-AD self and proxy scores were combined in the ratio 2 to 1</li> </ul>
			<ul style="list-style-type: none"> <li>Convenience sample of n = 17 people with dementia from the IC sample</li> <li>Inclusion (participants): People with dementia from the IC sample who were able to complete the QoI-AD 6 months after the first observation</li> </ul>	TRR	0.89 ( $\alpha$ )	?	<ul style="list-style-type: none"> <li>No sample size calculation</li> <li>Inappropriate use of Cronbach's alpha as statistical measure of agreement</li> <li>No information about blinding reported</li> <li>Second observation after 6 months</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Thorgrims et al. 2003 [76]	QoI-AD (GB) ▪ 13 items ▪ No subscales	Investigation of the validity and reliability of the QoI-AD as a scale for measuring QoI for people with dementia	<ul style="list-style-type: none"> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Convenience sample of n = 201 people with dementia recruited in residential homes and day centers</li> <li>▪ Inclusion (participants): Dementia diagnosis based on DSM-IV, MMSE between 10 and 24, ability to communicate and understand communication, ability to see and hear enough for the data collection, no behavior like constant wandering, shouting or aggression, no diagnosis of learning disability, no current clinical depression</li> <li>▪ Exclusion (participants): No ability to complete the 3 screening items of the DQoI</li> </ul>	IC	0.82	+	<ul style="list-style-type: none"> <li>▪ Analysis of Cronbach's alpha was based on the results of a principal component analysis (1 factor)</li> <li>▪ In contrast to the inclusion criteria people with MMSE score &gt; 24 were also included in the final sample</li> </ul>
			<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 38 people with dementia recruited in residential homes, nursing homes, day centers and hospitals</li> <li>▪ Inclusion (participants): Dementia diagnosis based on DSM-IV criteria</li> <li>▪ Exclusion (participants): No ability to complete the 3 screening items of the DQoI</li> </ul>	TRR	$\geq 0.60$	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ Only total ICC values without CI were reported, ICC values were computed for each items</li> <li>▪ No information about blinding reported</li> <li>▪ Second observation after 7 days</li> </ul>
Smith et al. 2005 [21]	QoI-AD (GB) ▪ 13 items ▪ No subscales	Psychometric evaluation of the DEMQoI and DEMQoI-proxy in a final field test (QoI-AD scores were compared to DEMQoI scores)	<ul style="list-style-type: none"> <li>▪ Random sample of n = 40 people with dementia (n = 55 caregivers) based on a convenience sample of n = 101 people with dementia (n = 99 caregivers) recruited from community mental health teams and Healthcare National Health Service Trust in the UK</li> <li>▪ Inclusion (people with dementia): dementia diagnosis based on ICD-10 criteria</li> <li>▪ No exclusion criteria</li> </ul>	IC	0.84  0.84 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Analysis of Cronbach's alpha was not based on the results of the conducted factor analysis</li> </ul>
Novelli et al. 2005 [65]	QoI-AD (BR) ▪ 13 items ▪ No subscales	Cross-cultural adaption and psychometric evaluation of the Portuguese-version (Brazilian) of the QoI-AD scale	<ul style="list-style-type: none"> <li>▪ Random sample of n = 40 people with dementia recruited from an outpatient clinic</li> <li>▪ Inclusion (participants): Alzheimer's disease based on NINCDS-ADRDA criteria with mild to moderate symptomatology according to DSM-III-R14 criteria, MMSE scores and judgment of their respective family caregivers</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	0.81  0.85 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
	QoI-AD (BR) ▪ 13 items ▪ No subscales		▪ Convenience sample of n = 17 people of the IC sample, all other Sample information was equal to the IC sample	TRR	0.71 <sup>e</sup>  1.00 <sup>e</sup> (proxy)	?	▪ No sample size calculation conducted ▪ No selection criteria for the TRR sample reported ▪ No characteristics for the TRR subsample of people with dementia and proxy-raters reported ▪ Only total Kappa values without CI were reported ▪ No information about blinding reported ▪ Second observation after 2 week
Novelli et al. 2010 [77]	QoI-AD (BR) ▪ 13 items ▪ No subscales	Evaluation of the reliability and construct validity of the QoI-AD in the Brazilian Version	▪ Convenience sample of n = 60 people with dementia recruited from an outpatient clinic ▪ Inclusion (participants): Probable Alzheimer's disease based on NINCDS-ADRDA criteria, DSM-III-R criteria, MMSE scores between 10 to 25, age $\geq$ 60 years, availability of a caregiver ▪ Exclusion (participants): Presence of serious language impairment  ▪ Inclusion (proxies): Patient family caregivers with $\geq$ 24 hours weekly contact time with the participants ▪ Exclusion (proxies): Not reported ▪ Proxy-training: Not reported	IC	0.80  0.83 (proxy)	?	▪ Small sample size ▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis
Fuh et al. 2006 [66]	QoI-AD (TW) ▪ 13 items ▪ No subscales	Examination of the psychometric properties of the QoI-AD in Taiwan	▪ Consecutive sample of n = 81 people with dementia recruited from an outpatient clinic ▪ Inclusion (participants): Dementia diagnosis based on the DSM-IV criteria ▪ Exclusion (participants): Not reported  ▪ Inclusion (proxies): Primary caregivers who lived with the participants or were in contact with them every day ▪ Exclusion (proxies): Not reported ▪ Proxy-training: Self- and proxy-ratings were guided by a research assistant	IC	0.83  0.79 (proxy)	?	▪ Small sample size ▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis ▪ No culture adaption reported

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Matsui et al. 2006 [67]	QoI-AD (JP) ▪ 13 items ▪ No subscales	Development of the Japanese version of the QoI-AD and investigation of its psychometric properties	<ul style="list-style-type: none"> <li>▪ Consecutive sample of n = 140 people with dementia recruited from 3 hospitals</li> <li>▪ Inclusion (participants): Diagnosis of probable Alzheimer's disease based on NINCDS-ADRDA criteria, residence with a caregiver in the community</li> <li>▪ Exclusion (participants): Neurological diseases other than Alzheimer's diseases, history of severe mental illnesses or substance abuse before the onset of dementia, MMSE &lt; 10, problems to understanding the questions</li> <li>▪ Inclusion (proxies): Caregivers of the participant</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	0.84  0.82 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>
			<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 83 people of the IC sample, all other Sample information was equal to the IC sample</li> </ul>	TRR	0.84 (0.76-0.91)  0.91 (0.86-0.94) (proxy)	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ No selection criteria for the TRR sample reported</li> <li>▪ No characteristics for the TRR subsample of people with dementia and proxy-raters reported</li> <li>▪ No information about blinding reported</li> <li>▪ Second observation after 1 month</li> </ul>
Tatsumi et al. 2011 [78]	QoI-AD (JP) ▪ 13 items ▪ No subscales	Examination of the applicability of the Japanese version of the QoI-AD for patients with mild cognitive impairment	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 47 people with mild cognitive impairment living at home</li> <li>▪ Inclusion (participants): Diagnosis of amnesic MCI based on the criteria of Petersen: (1) complaints of forgetfulness by the participant or by family members, (2) maintenance of independent everyday life activities, (3) no history of stroke or alcohol or drug dependence, (4) a CDR score of 0.5 and a MMSE <math>\geq</math> 25</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	0.90  0.93 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Lin Kiat Yap et al. 2008 [68]	QoI-AD (CN Mandarin) ▪ 13 items ▪ No subscales	Examination of the psychometric properties and clinical experience in using the Mandarin translation of the QoI-AD	<ul style="list-style-type: none"> <li>▪ Purposive sample of n = 67 people with dementia recruited in a geriatric outpatient clinic</li> <li>▪ Inclusion (participants): Diagnosis of dementia based on the DSM-III-R criteria, MMSE <math>\geq</math> 10, ability to give a response to more than 2 items of the QoI-AD</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Age <math>\geq</math> 18 years, conversant with English and Mandarin language, immediate family member and spending at least half a day in a week with the person with dementia</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Prior to QoI rating initial instructions by a trained research assistant</li> </ul>	IC	0.90  0.81 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>
				TRR	0.72  0.81 (proxy)	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ No selection criteria for the TRR sample reported</li> <li>▪ No characteristics for the TRR subsample of people with dementia and proxy-raters reported</li> <li>▪ Only total ICC values without CI were reported</li> <li>▪ No information about blinding reported</li> <li>▪ Second observation after 1 to 2 weeks</li> </ul>
Yu et al. 2013 [79]	QoI-AD (CN Mandarin) ▪ 13 items ▪ No subscales	Evaluation of the reliability and validity of the Chinese translation of the QoI-AD	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 195 people with dementia and 87 caregivers recruited in hospitals and communities</li> <li>▪ Inclusion (participants): Probable or possible Alzheimer's disease based on NINCDS-ADRDA criteria, actively involved caregiver who lived with the person with dementia or spent at least half a day in a week with them, community or ambulatory dwelling, age <math>\geq</math> 65 years, native Chinese speakers</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	0.66  0.87 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
			<ul style="list-style-type: none"> <li>Convenience sample of n= 50 people of the IC sample, all other sample information was equal to the IC sample</li> </ul>	TRR	0.84 0.90 (proxy)	?	<ul style="list-style-type: none"> <li>No sample size calculation conducted</li> <li>No selection criteria for the TRR sample reported</li> <li>No characteristics for the TRR subsample of people with dementia and proxy-raters reported</li> <li>Only total ICC values without CI were reported</li> <li>No information about blinding reported</li> <li>Second observation after 2 weeks</li> </ul>
Chan et al. 2011 [69]	QoI-AD (CN Cantonese) <ul style="list-style-type: none"> <li>13 items</li> <li>No subscales</li> </ul>	Investigation of the psychometric properties and effects of cognitive function as well as depressive mood on the QoI of Hong Kong Chinese Alzheimer's disease patients	<ul style="list-style-type: none"> <li>Convenience sample of n = 111 people with dementia recruited from 1 memory clinic and 1 psychogeriatric clinic</li> <li>Inclusion (participants): Chinese ethnicity, dementia diagnosis based on the DSM-IV criteria, Alzheimer's disease based on the NINCDS-ADRDA criteria, age <math>\geq</math> 65 years, living in the community, available caregiver who lived with the person with dementia</li> <li>Exclusion (participants): Age &lt; 65 years, severe visual or hearing impairment, living in a nursing home, inability to understand the items of the QoI-AD, psychiatric diagnoses including schizophrenia, alcoholic and substance abuse, cerebral infarction and mixed types of dementia</li> <li>Inclusion (proxies): Not reported</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Not reported</li> </ul>	IC	0.92 0.78 (proxy)	?	<ul style="list-style-type: none"> <li>Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>
			<ul style="list-style-type: none"> <li>Convenience sample of n = 30 people of the IC sample, all other sample information were equal to the IC sample</li> </ul>	TRR	0.96 0.97 (proxy)	?	<ul style="list-style-type: none"> <li>No sample size calculation conducted</li> <li>No selection criteria for the TRR sample reported</li> <li>No characteristics for the TRR subsample of people with dementia and proxy-raters reported</li> <li>Only total ICC values without CI were reported</li> <li>No information about blinding reported</li> <li>Length of period between first and second observation was not reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Wolak et al. 2009 [70]	QoI-AD (FR) ▪ 13 items ▪ No subscales	Evaluation of psychometric properties and the cross-cultural adaption into French of the quality of life in Alzheimer's disease scale	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 120 people with dementia recruited from 7 French hospitals</li> <li>▪ Inclusion (participants): Age <math>\geq</math> 65 years, native French speakers, living at home or in an institution, diagnosis of Alzheimer's disease based on the DSM IV criteria, MMSE score <math>\geq</math> 10, availability of a main caregiver who was able to read</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Ratings were guided by trained interviewers</li> </ul>	IC	0.83 0.79 (proxy) 0.84 (2*self and 1*proxy)	?	<ul style="list-style-type: none"> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>
				TRR	0.89 0.86 (proxy)	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ Only total ICC values without CI were reported</li> <li>▪ No information about blinding reported</li> <li>▪ Second observation after 2 weeks</li> </ul>
Wolak-Thierry et al. 2014 [57]	QoI-AD (FR) ▪ 13 items ▪ No subscales	Comparison of psychometric properties of the DQoI and QOL-AD to identify which instrument is most suitable for the use for consultations in geriatric medicine	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 123 people with Alzheimer's disease and n = 123 caregivers recruited in a geriatric department or a memory clinic</li> <li>▪ Inclusion (participants): Diagnosis based on DSM-IV and NINCDS-ADRDA criteria, age <math>\geq</math> 65 years, MMSE score <math>\geq</math> 10</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	0.83 0.77 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>
				TRR	0.80 0.70 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Only total ICC without CI was reported</li> <li>▪ No information about blinding reported</li> <li>▪ No sample size calculation</li> <li>▪ Second observation after 14 days</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Leon-Salas et al. 2011 [80]	QoI-AD (ES) ▪ 13 items ▪ No subscales	Evaluation of the psychometric properties of the Spanish version of the QoI-AD in institutionalized patients and families	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 40 people with AD and n = 101 family caregivers recruited from 1 long-term care facility</li> <li>▪ Inclusion (participants): Institutionalized people, diagnosis of Alzheimer's disease based on the NINCDS-ADRDA criteria or dementia mixed with AD</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Family members who identified themselves as the closest relative</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	0.90  0.86 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample size (n = 61 not completed by Residents)</li> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> <li>▪ No information about the cross-cultural adaption process reported</li> </ul>
				TRR	0.83 (0.66-0.92)  0.51 (0,17-0,75) (proxy)	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ No selection criteria for the TRR sample reported</li> <li>▪ No characteristics for the TRR subsample of people with dementia and proxy-raters reported</li> <li>▪ No information about blinding reported</li> <li>▪ Second observation after 3 weeks</li> </ul>
Gomez-Gallego et al. 2014 [71]	QoI-AD (ES) ▪ 13 items  Factor analysis: A: Health (4) B: Environment (4) C: Functional ability (5)	Investigation of the factorial structure of the QoI-AD in patients with Alzheimer's disease	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 139 people with dementia recruited from psychogeriatric clinics, day hospitals and day care centers</li> <li>▪ Inclusion (participants): Diagnosis of Alzheimer's disease based on the NINCDS-ADRDA criteria, GDS score between 3 to 5 and MMSE &gt; 11</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ A confirmatory factor analysis resulted in a 3 factor model</li> </ul>	IC	A: 0.79 B: 0.52 C: 0.71	+	

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Akpinar et al. 2012 [72]	QoI-AD (TR) ▪ 13 items ▪ No subscales	Investigation of the validity and reliability of the Turkish version of the QoI-AD	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 72 people with dementia living in the community</li> <li>▪ Inclusion (participants): Dementia diagnosis based on the DSM-IV criteria, MMSE <math>\geq</math> 10</li> <li>▪ Exclusion (participants): People with behavioral problems impeding the data collection (e. g. screaming, wandering), people with visual or hearing impairment</li> <li>▪ Inclusion (proxies): Primary caregiver who provide care at least 24 hours a week</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: QoI ratings were guided by a researcher, in cases of misunderstanding the researcher gave explanations</li> </ul>	IC	0.84  0.77 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>
Barríos et al. 2013 [73]	QoI-AD (PT) ▪ 13 items ▪ No subscales	Translation, adaption and validation of the QoI-AD to European Portuguese	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 104 people with dementia recruited from 1 hospital and 1 long-term care unit</li> <li>▪ Inclusion (participants): Age <math>\geq</math> 50 years, native Portuguese speakers, diagnosis of mild cognitive impairment based on criteria of the European Consortium on Alzheimer's disease, Alzheimer's disease based on the NINCDS-ADRDA criteria, diagnosis of vascular dementia, MMSE <math>\geq</math> 10</li> <li>▪ Exclusion (participants): Acute health problems in the 30 days prior to the data collection, severe visual or hearing impairment</li> <li>▪ Inclusion (proxies): Caregivers responsible for supervision and care of the participants, <math>\geq</math> 20 hour per week contact with the patient, able to read and write, when no family caregiver was available a professional caregiver who met the inclusion criteria was included as proxy.</li> <li>▪ Exclusion (proxies): Caregivers with cognitive deficits</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	0.87  0.86 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Analysis of Cronbach's alpha was not based on the results of a factor analysis</li> </ul>
			<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 20 people of the IC sample settled in a long-term care unit, all other sample information was equal to the IC sample</li> </ul>	TRR	0.92 (0.80-0.97)  0.92 (0.82-0.97) (proxy)	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ No characteristics for the TRR subsample of people with dementia and proxy-raters reported</li> <li>▪ No information about blinding reported</li> <li>▪ Second observation after 2 weeks</li> </ul>
Graske et al. 2014	QoI-AD (DE) ▪ 13 items	Evaluation of psychometric properties	<ul style="list-style-type: none"> <li>▪ Convenience sample recruited in 36 shared housing arrangements in Berlin resulting in a sample of n = 97</li> </ul>	IC	0.74 (proxy)	+	<ul style="list-style-type: none"> <li>▪ No information about cross-cultural adaptation reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
[81]	<ul style="list-style-type: none"> <li>No subscales</li> </ul>	of 3 QoI instruments in German shared housing arrangements	people with cognitive impairment or dementia <ul style="list-style-type: none"> <li>Inclusion (participants): Not reported</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (proxies): Registered nurses working in the recruited shared housing arrangements</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Nurses received a training session prior to data collection including carrying out interviews with people with cognitive impairment</li> </ul>				<ul style="list-style-type: none"> <li>Unidimensionality of the QoI-AD was confirmed by a confirmatory factor analysis</li> <li>n = 36 assessments with 1 or 2 missing were imputed</li> </ul>
<b>Quality of Life in Alzheimer's Disease Nursing home version, QoI-AD NH (proxy)</b>							
Edelman et al. 2014 [52]	QoI-AD NH (US) <ul style="list-style-type: none"> <li>15 items</li> <li>No subscales</li> </ul>	Comparison of the properties of different dementia-specific QoI measurements	<ul style="list-style-type: none"> <li>Convenience sample of n = 54 people with dementia for the proxy rating and n = 36 people with dementia for the self-rating, recruited from 3 adult day care centres</li> <li>Inclusion (participants): MMSE <math>\geq</math> 10</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (proxies): Staff members who were highly familiar with the particular participants</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Not reported</li> </ul>	IC	0.92 0.88 (proxy)	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>Results for internal consistency were not based on the results of a factor analysis</li> <li>No characteristics for the sample of proxy-raters reported</li> </ul>
Edelman et al. 2005 [35]	QoI-AD NH (US) <ul style="list-style-type: none"> <li>15 items</li> <li>No subscales</li> </ul> Original version (US)	Investigation of the distribution and factor structure of the items and psychometric properties of the dementia-specific QoI instrument	<ul style="list-style-type: none"> <li>Convenience sample of n = 168 people with dementia for the proxy rating and n = 65 people with dementia for the self-rating, recruited from 4 dementia-specific nursing homes and 3 assisted living facilities</li> <li>Inclusion (participants): MMSE <math>\geq</math> 10</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (proxies): Staff member who were most familiar with the particular clients</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Not reported</li> </ul>	IC	0.94 0.90 (proxy)	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>Evaluation of internal consistency was not based on the results of a conducted exploratory factor analysis</li> <li>No characteristics for the sample of proxy-raters reported</li> </ul>
Sloane et al. 2005 [53]	QoI-AD NH (US) <ul style="list-style-type: none"> <li>15 items</li> <li>No subscales</li> </ul>	Investigation of psychometric properties of QoI measures in long-term care facilities	<ul style="list-style-type: none"> <li>Purposive sample of n = 410 people with dementia for the proxy rating and n = 121 people with dementia for the self-rating, recruited from 45 residential care facilities, assisted living facilities and nursing homes</li> <li>Inclusion (participants): Age <math>\geq</math> 65 years, diagnosis of dementia</li> <li>Exclusion (participants): Primary diagnosis of Huntington's</li> </ul>	IC	0.92 0.88 (proxy)	?	<ul style="list-style-type: none"> <li>No factor analysis conducted</li> <li>No characteristics for the subsample of self-rating people with dementia and proxy-raters reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
			disease, alcohol related dementia, schizophrenia, manic-depressive disorders, mental retardation, MMSE $\geq$ 10  <ul style="list-style-type: none"> <li>▪ Inclusion (proxies): Registered nurses or nursing assistant who know the resident best</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>				
			<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 20 people of the IC sample, all other sample information was equal to the IC sample</li> </ul>	IRR	0.99 (proxy)	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ No selection criteria for the IRR sample reported</li> <li>▪ No characteristics for the sample of people with dementia and proxy-raters reported</li> <li>▪ Only total ICC values without CI were reported</li> <li>▪ No information about blinding reported</li> </ul>
Moyle et al. 2012 [51]	QoI-AD NH (US) <ul style="list-style-type: none"> <li>▪ 15 items</li> <li>▪ No subscales</li> </ul>	Comparative analysis of the DQoI and QoI-AD-self NH as QoI instruments in nursing homes	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 60 people with dementia recruited from 4 long-term care facilities</li> <li>▪ Inclusion (participants): Confirmed diagnosis of dementia or probable dementia based on a MMSE score between 10 to 24 or features consistent with the DSM-IV criteria, ability to complete surveys with the aid of an assistant, one family member or friend who was accessible and in contact with the participant</li> <li>▪ Exclusion (participants): Not reported</li> </ul>	IC	0.86	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Evaluation of internal consistency was not based on a factor analysis</li> <li>▪ US Version used in an AU setting</li> </ul>
<b>Quality of Life Assessment Schedule, QOLAS (proxy)</b>							
Selai et al. 2001 [36]	QOLAS (GB) <ul style="list-style-type: none"> <li>▪ 10 items</li> <li>A: Physical (2)</li> <li>B: Psychological (2)</li> <li>C: Social/Family (2)</li> <li>D: Usual activities (2)</li> <li>E: Cognitive functioning (2)</li> </ul>	Adaptation and psychometric evaluation of a generic QoI instrument to a dementia-specific QoI instrument	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 22 people with dementia recruited from the outpatient clinics of a hospital</li> <li>▪ Inclusion (participants): Ability to give meaningful replies during the QOLAS interview</li> <li>▪ Exclusion (participants): Lack of regular caregiver</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	0.78  0.78 (proxy)	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Evaluation of internal consistency was not based on a factor analysis</li> </ul>
Original version (GB)							
<b>Proxy-rated only</b>							

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
<b>Alzheimer Disease Related Quality of Life, ADRQL (Japanese version called AD-HRQL-J or later QLDJ)</b>							
Edelman et al. 2004 [52]	ADRQL (US) ▪ 47 items  Original subscales: A: Social interaction (12) B: Awareness of self (8) C: Feelings and mood (15) D: Enjoyment of activities (5) E: Response to surroundings (7)	Comparison of the properties of different dementia-specific QoI measurements	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 54 people with dementia recruited from 3 adult day care centres</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Staff persons who were highly familiar with the particular clients</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	0.86	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Results for internal consistency were not based on the results of a factor analysis</li> <li>▪ Doubtful calculation of Cronbach's alpha according to the multidimensionality of the ADRQL</li> </ul>
Edelman et al. 2005 [35]	ADRQL (US) ▪ 47 items ▪ Original subscales	Investigation of the distribution and factor structure of the items and psychometric properties of the dementia-specific QoI instrument	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 172 people with dementia recruited from 4 dementia-specific nursing homes and 3 assisted living facilities</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Staff member who were most familiar with the particular clients</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	A: 0.80 B: 0.68 C: 0.71 D: 0.61 E: 0.29	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Results for internal consistency were not based on the results of a conducted exploratory factor analysis</li> <li>▪ No characteristics for the sample of proxy-raters reported</li> </ul>
Sloane et al. 2005 [53]	ADRQL (US) ▪ 47 items ▪ Original subscales	Investigation of psychometric properties of QoI measures in long-term care facilities	<ul style="list-style-type: none"> <li>▪ Purposive sample of n = 414 people with dementia recruited from 45 residential care facilities, assisted living facilities and nursing homes</li> <li>▪ Inclusion (participants): Age <math>\geq</math> 65 years, diagnosis of dementia</li> <li>▪ Exclusion (participants): Primary diagnosis of Huntington's disease, alcohol related dementia, schizophrenia, manic-depressive disorders, mental retardation</li> <li>▪ Inclusion (proxies): Registered nurses or nursing assistant who know the resident best</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	A: 0.69 B: 0.62 C: 0.72 D: 0.50 E: 0.46	?	<ul style="list-style-type: none"> <li>▪ Evaluation of internal consistency was not based on a factor analysis</li> <li>▪ No characteristics for the proxy-raters reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Sloane et al. 2005 [53]	ADRQL (US) ▪ 47 items ▪ Original subscales		▪ Convenience sample of n = 23 people of the IC sample, all other sample information was equal to the IC sample	IRR	A: 1.00 B: 0.90 C: 0.98 D: 0.97 E: 1.00	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ No selection criteria for the IRR sample reported</li> <li>▪ No characteristics for the sample of people with dementia and proxy-raters reported</li> <li>▪ Only total ICC values without CI were reported</li> <li>▪ No information about blinding reported</li> </ul>
Kasper et al. 2009 [85]	ADRQL (US) ▪ 47 items ▪ Original subscales  ADRQL revised (US) ▪ [40 items]  A: Social interaction [12] B: Awareness of self [8] C: Feelings and mood [12] D: Enjoyment of activities [4] E: Response to surroundings [4]	Examination of various psychometric properties of the ADRQL	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 369 people with dementia recruited from 3 primary samples</li> <li>▪ People were recruited from the community n = 146, nursing homes n = 89 and assisted living facilities n = 134</li> <li>▪ Inclusion (participants): Diagnosis of Alzheimer's disease based on ICD-9 criteria (community), AD diagnoses in medical records (other settings)</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Close family members or professional caregivers</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	ADRQL (US): A: 0.83 B: 0.79 C: 0.81 D: 0.58 E: 0.56  ADRQL revised (US): A: 0.83 B: 0.79 C: 0.81 D: 0.61 E: 0.56	?	<ul style="list-style-type: none"> <li>▪ Evaluation of internal consistency was not based on a factor analysis</li> <li>▪ No characteristics for the sample of proxy-raters reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Yamamoto et al. 2000 [84]	ADRQL (JP) called AD-HRQL-J ▪ 48 items  A: Social interaction (12) B: Awareness of self (8) C: Feelings and mood (15) D: Enjoyment of activities (6) E: Response to surroundings (7)	Examination of the psychometric properties of the Japanese translation of the ADRQL	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 296 participants recruited from health care facilities Japan</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li> </ul>	IC	A: 0.67 B: 0.88 C: 0.51 D: 0.88 E: 0.87	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Evaluation of internal consistency was not based on the results of a factor analysis</li> <li>▪ No characteristics for the sample of proxy-raters reported</li> <li>▪ Participants were not necessarily people with dementia</li> </ul>
			<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 25 people with dementia recruited from long-term care facilities Japan</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IRR	Ranges: A: -0.24-0.60 <sup>e</sup> B: -0.06-0.43 C: -0.09-1.00 D: 0.04-0.58 E: 0.02-0.50	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ No characteristics for the sample of proxy-raters reported</li> <li>▪ Participants were not necessarily people with dementia</li> <li>▪ Only total Kappa values without CI were reported, Kappa values were calculated for each item</li> <li>▪ No information about blinding reported</li> </ul>
			<ul style="list-style-type: none"> <li>▪ Inclusion (proxies): Not reported</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	TRR	Ranges: A: -0.07-0.90 <sup>e</sup> B: 0.35-1.00 C: 0.43-1.00 D: 0.39-0.69 E: 0.00-0.80	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation conducted</li> <li>▪ No characteristics for the sample of proxy-raters reported</li> <li>▪ Participants were not necessarily people with dementia</li> <li>▪ Only total kappa values without CI were reported, kappa values were calculated for each item</li> <li>▪ No information about blinding reported</li> <li>▪ Second observation after 14 days</li> </ul>
Yamamoto-Mitani et al. 2002 [86]	ADRQL (JP) called QLDJ ▪ 24 items  A: Interacting with surroundings (8)	Development and psychometric evaluation of the QLDJ	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 623 people with dementia recruited from health care facilities</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Nurses or social workers, Close</li> </ul>	IC	A: 0.89 B: 0.88 C: 0.80	+	<ul style="list-style-type: none"> <li>▪ Evaluation of internal consistency was based on the results of a principal component analysis</li> <li>▪ No characteristics for the sample of proxy-raters reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
	B: Expressing self (10) C: Experiencing minimum negative behavior (6)		relationship to people with dementia <ul style="list-style-type: none"> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Not reported</li> </ul>				
			<ul style="list-style-type: none"> <li>Convenience sample of n = 36 people with dementia recruited from health care facilities</li> <li>People with dementia were rated between four and five times from 16 proxy-raters resulting in paired 160 QoI-ratings in total</li> <li>Inclusion (participants): Not reported</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (proxies): Nurses or social workers</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Not reported</li> </ul>	IRR	A: 0.66 <sup>f</sup> B: 0.71 C: 0.67	?	<ul style="list-style-type: none"> <li>No sample size calculation conducted</li> <li>No characteristics for the sample of proxy-raters reported</li> <li>Only total weighted Kappa values without CI was reported</li> <li>No information about blinding reported</li> </ul>
			<ul style="list-style-type: none"> <li>Sample information was equal to the IRR sample</li> <li>People with dementia were rated between five and eight times from 19 proxy-raters resulting in 115 paired QoI-ratings in total</li> </ul>	TRR	A: 0.89 <sup>f</sup> B: 0.91 C: 0.86	?	<ul style="list-style-type: none"> <li>No sample size calculation conducted</li> <li>No characteristics for the sample of proxy-raters reported</li> <li>Only total weighted kappa values without CI was reported</li> <li>No information about blinding reported</li> <li>Second observation after 14 days</li> </ul>
Menzi-Kuhn 2006 [83]	ADRQL (CH) <ul style="list-style-type: none"> <li>47 items</li> <li>A: Social interaction (12)</li> <li>B: Awareness of self (8)</li> <li>C: Feelings and mood (15)</li> <li>D: Enjoyment of activities (5)</li> <li>E: Response to surroundings (7)</li> </ul>	Investigation of the Quality of life of people with dementia living in nursing homes in Switzerland	<ul style="list-style-type: none"> <li>Convenience sample of n = 95 people with dementia recruited from 6 nursing homes</li> <li>Inclusion (participants): Dementia diagnosis confirmed by a physician, age <math>\geq 65</math> years, residence in the nursing home <math>\geq 2</math> months</li> <li>Exclusion (participants): very severe dementia combined with bedridden, loss of communication abilities and comprehensive care dependency, severe additional diseases like oncological diseases, blindness, hearing impairment</li> <li>Inclusion (proxies): Nurses (n = 26) with a close relationship to the particular participant, working contract for at least half-time work, working in the nursing home for <math>\geq 3</math> months, and at work in the two weeks prior to the QoI rating</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Proxy-based QoI rating was guided by a researcher</li> </ul>	IC	A: 0.59 B: 0.62 C: 0.59 D: 0.39 E: 0.47	-	<ul style="list-style-type: none"> <li>Evaluation of internal consistency was not based on a factor analysis</li> <li>Results are based on Swiss item weights which were different to the item weights of the original US-version</li> </ul>
			<ul style="list-style-type: none"> <li>Convenience sample of n = 18 people with dementia recruited from 4 nursing homes. All other sample</li> </ul>	IRR	A: 0.36 B: 0.61	?	<ul style="list-style-type: none"> <li>No sample size calculation conducted</li> <li>No selection criteria for the IRR sample</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
			information was equal to the IC sample (n = 12 staff)		C: 0.13 D: 0.02 E: -0.17		reported <ul style="list-style-type: none"> <li>No characteristics for the IRR subsample of people with dementia and proxy-raters reported</li> <li>Only total ICC values without CI were reported</li> <li>No information about blinding reported</li> </ul>
Graske et al. 2014 [81]	ADRQL (DE) <ul style="list-style-type: none"> <li>47 items</li> <li>A: Social interaction (12)</li> <li>B: Awareness of self (8)</li> <li>C: Feelings and mood (15)</li> <li>D: Enjoyment of activities (5)</li> <li>E: Response to surroundings (7)</li> </ul>	Evaluation of psychometric properties of 3 QoI instruments in German shared housing arrangements	<ul style="list-style-type: none"> <li>Convenience sample recruited in 36 shared housing arrangements in Berlin resulting in a sample of n = 104 people with cognitive impairment or dementia</li> <li>Inclusion (participants): Not reported</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (proxies): Registered nurses working in the recruited shared housing arrangements</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Nurses received a training session prior to data collection including carrying out interviews with people with cognitive impairment</li> </ul>	IC	A: 0.72 B: 0.54 C: 0.77 D: 0.44 E: 0.45	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>No information about cross-cultural adaptation reported</li> <li>Subscales were confirmed by a confirmatory factor analysis</li> </ul>
<b>Dementia Care Mapping, DCM</b>							
Fossey et al. 2002 [87]	DCM 7 <sup>th</sup> Edition (GB) <ul style="list-style-type: none"> <li>26 items</li> <li>Original subscales: A: Well being</li> <li>B: Activity</li> </ul>	Examination of the internal consistency, test-retest reliability and concurrent validity of Dementia Care Mapping	<ul style="list-style-type: none"> <li>Random sample of at least 50% (n = 123) of residents with dementia recruited in 6 nursing homes</li> <li>Inclusion (participants): Not reported</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (observers): Not reported</li> <li>Exclusion (observers): Not reported</li> <li>Observer-training: 3 day licensed course with a formal assessment</li> </ul>	IC	---	0	<ul style="list-style-type: none"> <li>Correlations between the 3 main measures, activities, well- or ill-being score and social withdrawal were calculated. Despite the significant correlation this provides no information about the internal consistency of the DCM</li> </ul>
			<ul style="list-style-type: none"> <li>Convenience sample of n = 54 people with dementia recruited from 2 primary studies in nursing homes</li> <li>Inclusion (participants): Study 1: severe dementia (CDR = 3) and a clinical significant agitation. Study 2: not reported</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (observers): Not reported</li> <li>Exclusion (observers): Not reported</li> <li>Observer-training: Not reported</li> </ul>	TRR	A: 0.58 <sup>d</sup> B: 0.40	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>No characteristics of observers reported</li> <li>Only total Spearman rank correlation coefficients without CI were reported</li> <li>No information about blinding reported</li> <li>Second observation after 1-4 weeks</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Thornton et al. 2004 [88]	DCM 7 <sup>th</sup> Edition (GB) ▪ 26 items ▪ Original subscales	Evaluation of the psychometric properties of Dementia Care Mapping	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 20 people with dementia recruited from a NHS continuing care ward or a day hospital</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (observers): Not reported</li> <li>▪ Exclusion (observers): Not reported</li> <li>▪ Observer-training: Trained in using the DCM method</li> </ul>	IRR	Ranges: A: -0.003-0.52 <sup>e</sup> B: -0.004-1.00 <sup>e</sup>	-	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ No characteristics of people with dementia and observers reported</li> <li>▪ Only total Kappa values without CI were reported</li> <li>▪ Observer informed about a test situation, no other information about blinding reported</li> </ul>
Rüsing 2004 [89]	DCM 7 <sup>th</sup> Edition (DE) ▪ 26 items ▪ Original subscales	Evaluation of the inter-rater reliability of Dementia Care Mapping	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 38 people with dementia recruited from nursing homes</li> <li>▪ Inclusion (participants): Dementia diagnosis based on physician or registered nurse judgement, residence in a nursing home, ability to spend time in the public area of the nursing unit</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (observers): DCM basic user (n = 11)</li> <li>▪ Exclusion (observers): Not reported</li> <li>▪ Observer-training: Trained DCM basic user or DCM advanced user</li> </ul>	IRR	Ranges: A: 0.00-0.54 <sup>e</sup> B: 0.18-1.00 <sup>e</sup>	-	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Observers were not allowed to talk to each other during observation</li> </ul>
Sloane et al. 2005 [53]	DCM 7 <sup>th</sup> Edition (GB) ▪ 26 items ▪ Original subscales	Investigation of psychometric properties of QoI measures in long-term care facilities	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 56 people with dementia recruited from 7 nursing homes</li> <li>▪ Inclusion (participants): Dementia diagnosis, Age <math>\geq</math> 65 years</li> <li>▪ Exclusion (participants): Primary diagnosis of Huntington's disease, alcohol related dementia, schizophrenia, manic-depressive disorders, mental retardation</li> <li>▪ Inclusion (observers): Not reported</li> <li>▪ Exclusion (observers): Not reported</li> <li>▪ Observer-training: Not reported</li> </ul>	IRR	A: 0.70 B: 0.80 (DCM Type I)	?	<ul style="list-style-type: none"> <li>▪ No sample size calculation</li> <li>▪ The original GB-version of the DCM was used in a US sample without cross-cultural adaptation</li> <li>▪ No characteristics for the sample of people with dementia and proxy-raters reported</li> <li>▪ Only total ICC values without CI were reported</li> </ul>
Sloane et al. 2007 [90]	DCM 7 <sup>th</sup> Edition (GB) ▪ 26 items ▪ Original subscales	Investigation of item distribution, inter-rater reliability, validity and alternative data collection and coding methods for DCM based on a secondary data analysis of multiple prior studies	<ul style="list-style-type: none"> <li>▪ Convenience sample of people with dementia recruited from long-term care facilities</li> <li>▪ Inclusion (participants): Dementia diagnosis, Age <math>\geq</math> 65 years</li> <li>▪ Exclusion (participants): Primary diagnosis of Huntington's disease, alcohol related dementia, schizophrenia, manic-depressive disorders, mental retardation</li> <li>▪ Inclusion (observers): One highly and one less experienced DCM mapper</li> </ul>	IRR	A: 0.32 <sup>a</sup> B: 0.54 <sup>e</sup>	-	<ul style="list-style-type: none"> <li>▪ No sample size calculation</li> <li>▪ No information about cross-cultural adaptation reported</li> <li>▪ No information about the exact number of participating people with dementia</li> <li>▪ No characteristics for the sample of people with dementia and proxy-raters reported</li> <li>▪ Only total Pearson correlation</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
			<ul style="list-style-type: none"> <li>Exclusion (observers): Not reported</li> <li>Observer-training: Not reported</li> </ul>				<ul style="list-style-type: none"> <li>coefficients and kappa values without CI were reported</li> <li>No information about blinding reported</li> </ul>
<b>Observing Quality of Life in Dementia, OQOLD (Observed Quality of Life in Dementia Advanced, OQOLDA)</b>							
Edelman et al. 2007 [40]	OQOD (OQOLDA), (US) <ul style="list-style-type: none"> <li>6 items for Observation</li> <li>No subscale</li> </ul> Original version (US)	Development and evaluation of two new observational dementia-specific quality of life measure (OQOLD & QOLDA)	<ul style="list-style-type: none"> <li>Convenience sample of n = 30 people with dementia recruited from a care facility</li> <li>Inclusion (participants): MMSE &lt; 10, FAST score = 7a-e</li> <li>Exclusion (participants): No criteria reported</li> <li>Inclusion (observers): Nurses (n = 5)</li> <li>Exclusion (observers): Not reported</li> <li>Observer-training: Not reported</li> </ul>	IRR	OQOLDA scores: 0.87 (most observed)  0.95 (least observed)	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>No characteristics of observers reported</li> <li>Only total ICCs value without CI were reported</li> <li>No information about blinding reported</li> </ul>
<b>Psychological well-being in cognitively impaired persons, PWP-CIP</b>							
Burgener et al. 2005 [41]	PWP-CIP (US) <ul style="list-style-type: none"> <li>11 items</li> </ul> Factor analysis: <ul style="list-style-type: none"> <li>A: Positive affect/interaction (3)</li> <li>B: Negative affect/interaction (8)</li> </ul> Original version (US)	Investigation of the psychometric properties of the PWP-CIP instrument when used on persons with dementia as they progress from the early to middle disease stages	<ul style="list-style-type: none"> <li>Convenience sample of n = 95 (Follow-up n = 73) people with dementia recruited from 3 large diagnostic centres, people with dementia lived in private homes or assisted living arrangements</li> <li>Inclusion (participants): Confirmed dementia diagnosis (Alzheimer's disease, vascular dementia, Lewy body or mixed type dementia) within the last year prior to the study</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (proxies): Primary family caregivers</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Not reported</li> </ul>	IC	Baseline: A: 0.79 B: 0.74  Follow-Up (18 month): A: 0.84 B: 0.74	+	<ul style="list-style-type: none"> <li>Factor analysis confirmed the 2 subscales</li> </ul>
<b>Quality of Life for Dementia, QoI-D</b>							
Terada et al. 2002 [42]	QoI-D (JP) <ul style="list-style-type: none"> <li>31 items</li> </ul> Factor analysis: <ul style="list-style-type: none"> <li>A: Positive affect (7)</li> <li>B: Negative affect and actions (6)</li> <li>C: Ability of communication (5)</li> <li>D: Restlessness (5)</li> <li>E: Attachment with others (4)</li> </ul>	Development of the Quality of life questionnaire for dementia (QoI-D) in Japan	<ul style="list-style-type: none"> <li>Convenience sample of n= 264 people with dementia recruited from 2 psychiatric facilities, 1 hospital for the elderly with dementia, 1 nursing home, 3 geriatric health service facilities</li> <li>Inclusion (participants): Dementia diagnosis</li> <li>Exclusion (participants): Persons in bedridden state</li> <li>Inclusion (proxies): Nursing staff or family members who were well acquainted with the particular person with dementia</li> <li>Exclusion (proxies): No attempt was made to select raters</li> <li>Proxy-Training: Proxy-raters received a brief introduction to the QoI questionnaire, no training</li> </ul>	IC	A: 0.91 B: 0.88 C: 0.88 D: 0.79 E: 0.88 F: 0.83	+	<ul style="list-style-type: none"> <li>Factor analysis confirmed the 6 subscales</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
	F: Spontaneity and activity (4)  Original version (JP)		<ul style="list-style-type: none"> <li>Subsample of n = 10 people with dementia from the IC sample, selection criteria of the IRR subsample were not reported</li> <li>QoI of each person with dementia was assessed by four different proxy-raters</li> </ul>	IRR	A: 0.79 B: 0.90 C: 0.85 D: 0.76 E: 0.93 F: 0.63	?	<ul style="list-style-type: none"> <li>No sample size calculation</li> <li>No characteristics and selection criteria for the IRR subsample of people with dementia and proxy raters reported</li> <li>Only total average ICC value without CI were reported</li> <li>No information about blinding reported</li> </ul>
<b>Quality of Life in Late-stage Dementia Scale, QUALID</b>							
Weiner et al. 2000 [43]	QUALID (US) 11 items No subscales  Original version (US)	Development and psychometric evaluation of the QUALID instrument	<ul style="list-style-type: none"> <li>Convenience sample of n = 31 people with dementia recruited from dementia special care units</li> <li>Inclusion (participants): Not reported</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (proxies): Licensed vocational nurses and nursing assistants working on the recruited dementia special care units, <math>\geq 30</math> hours of exposure to the particular person with dementia in the week prior to data collection</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Proxy rating nurses were guided by a study nurse</li> </ul>	IC	0.77	?	<ul style="list-style-type: none"> <li>Small sample size</li> <li>No number and characteristics for the proxy-raters reported</li> <li>Principal component analysis confirmed the unidimensionality</li> </ul>
			<ul style="list-style-type: none"> <li>Subsample of n = 23 people with dementia from the IC sample, selection criteria of the IRR subsample were not reported</li> <li>Data were based on the ratings of n = 2 proxy rating nurses</li> <li>All other sample information was equal to the IC sample</li> </ul>	IRR	0.83 (0.07 <sup>b</sup> )	?	<ul style="list-style-type: none"> <li>No sample size calculation conducted</li> <li>No characteristics and selection criteria for the IRR subsample of people with mild to severe dementia and the proxy-raters reported</li> <li>No information about blinding reported</li> </ul>
			<ul style="list-style-type: none"> <li>Random subsample of n = 19 people with dementia from the IC sample</li> <li>All other sample information was equal to the IC sample</li> </ul>	TRR	0.81 (0.08 <sup>b</sup> )	?	<ul style="list-style-type: none"> <li>No sample size calculation conducted</li> <li>No characteristics and selection criteria for the TRR subsample of people with mild to severe dementia and the proxy raters reported</li> <li>No information about blinding reported</li> <li>Second observation after 2 to 3 days</li> </ul>
Falk et al. 2007 [91]	QUALID (SE) 11 items No subscales	Evaluation of the validity, reliability and responsiveness of a Swedish version of the QUALID	<ul style="list-style-type: none"> <li>Convenience sample of n = 169 people with dementia recruited from dementia special care units in 8 long-term care facilities</li> <li>Inclusion (participants): Not reported</li> <li>Exclusion (participants): Not reported</li> <li>Inclusion (proxies): Assistant nurses and untrained caregivers</li> </ul>	IC	0.74	+	<ul style="list-style-type: none"> <li>Principal component analysis confirmed the unidimensionality</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
			<ul style="list-style-type: none"> <li>who had extensive knowledge about the particular person with dementia and were responsible for the nursing care</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Not reported</li> </ul>				
			<ul style="list-style-type: none"> <li>Subsample of n = 26 people with dementia from the IC sample</li> <li>All other sample information was equal to the IC sample</li> </ul>	IRR	0.69	-	<ul style="list-style-type: none"> <li>No sample size calculation conducted</li> <li>No characteristics for the IRR subsample of people with dementia and the proxy raters reported</li> <li>No information about blinding reported</li> <li>Only total ICC value without CI were reported</li> </ul>
			<ul style="list-style-type: none"> <li>Subsample of n = 23 people with dementia from the IC sample, selection criteria of the TRR subsample were not reported</li> <li>Inclusion (participants): people with dementia who were judged to be unchanged in a global assessment of wellbeing based on a visual analogue scale</li> <li>All other Sample information was equal to the IC sample</li> </ul>	TRR	0.86	?	<ul style="list-style-type: none"> <li>No sample size calculation conducted</li> <li>No characteristics for the TRR subsample of people with dementia and the proxy-raters reported</li> <li>No information about blinding reported</li> <li>Only total ICC value without CI were reported</li> <li>Second observation after 3 weeks</li> </ul>
Garre-Olmo et al. 2010 [92]	QUALID (ES) <ul style="list-style-type: none"> <li>11 items</li> </ul> PCA analysis: <ul style="list-style-type: none"> <li>A: Signs of discomfort (4)</li> <li>B: Social interaction (4)</li> <li>C: Negative affective mood (3)</li> </ul>	Development and evaluation of the psychometric properties of the Spanish-version of the QUALID	<ul style="list-style-type: none"> <li>Convenience sample of 8 long-term care facilities, in each facility a random sample of 20 persons with dementia was performed (total n = 160)</li> <li>QoI of each person with dementia was assessed twice by one nurse from the morning shift and one from the afternoon shift</li> <li>Inclusion (participants): Severe dementia defined by an GDS score <math>\geq 6</math></li> <li>Exclusion (participants): Diagnosis of mental retardation or Schizophrenia</li> <li>Inclusion (proxies): Nurses from each long-term care facility who were responsible for the care of the participating people with dementia, one from the morning shift (n = 78) and one from the afternoon shift (n = 74)</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: Proxy rating nurses were guided by a trained interviewer</li> </ul>	IC	A: 0.71 B: 0.70 C: 0.68	+	<ul style="list-style-type: none"> <li>A principal component analysis identified 3 factors</li> </ul>
			<ul style="list-style-type: none"> <li>For information about the IRR sample n = 160 and proxy-raters see IC above</li> </ul>	IRR	0.74 (0.64-0.81)	+	

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Mjorud et al. 2014 [93]	QUALID (NO) ▪ 11 items  Factor analysis: A: Tension (5) B: Well-being (4) C: Sadness (3)	Investigation of the validity and internal consistency of the Norwegian QUALID version	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 661 people with dementia from 16 nursing homes</li> <li>▪ Inclusion (participants): Status of permanent nursing home resident, residence duration <math>\geq</math> 4 weeks, CDR score <math>\geq</math> 1, nonterminal illness</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Primary nurses (registered or auxiliary nurse)</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Proxy rating nurses were guided by a trained research assistant, leader of the study available during the data collection</li> </ul>	IC	A: 0.69 B: 0.62 C: 0.65	+	<ul style="list-style-type: none"> <li>▪ No information about cross-cultural adaptation reported</li> <li>▪ No characteristics of proxy-raters reported</li> <li>▪ One item included in two factors</li> </ul>
<b>QUALIDEM</b>							
Ettema et al. 2007 [10]	QUALIDEM (NL) ▪ 37 items ▪ [18 items for very severe dementia]  Mokken analysis: A: Care relationship (7), [3] B: Positive affect (6), [4] C: Negative affect (3), [2] D: Restless tense behavior (3), [3] E: Positive self-image (3) F: Social relations (6), [3] G: Social isolation (3), [3] H: Feeling at home (4) I: Having something to do (2)	Development of the QUALIDEM measurement and evaluation of its psychometric properties	<ul style="list-style-type: none"> <li>▪ Convenience sample of n = 238 people with dementia and n = 127 certified nursing assistants recruited from 25 psychogeriatric wards in 10 nursing homes</li> <li>▪ The total sample was divided in one subsample of people with mild to severe dementia (n = 202, GDS 2-6) and the total sample of people with mild to very severe dementia [n = 238, GDS 2-7]</li> <li>▪ Inclusion (participants): Age <math>\geq</math> 65 years, dementia diagnosis</li> <li>▪ Exclusion (participants): Diagnosis of Parkinson's disease, a diminution of consciousness or terminal care judged by the nursing home physician</li> <li>▪ Inclusion (proxies): Certified nursing assistants (CNA) working in the psychogeriatric wards</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: CNAs received 1 hour's oral instruction at the start of the observation period, a written instruction was available</li> <li>▪ Random sample of 7 people with dementia from each nursing home of the IC-sample resulting in a subsample of people with mild to severe dementia (n = 62) and the total sample of people with mild to very severe dementia [n=68]</li> <li>▪ For information about in- and exclusion criteria of people with dementia and proxy-raters see IC above</li> </ul>	IC	A: 0.83 [0.69] B: 0.89 [0.83] C: 0.71 [0.50] D: 0.74 [0.71] E: 0.64 F: 0.80 [0.59] G: 0.59 [0.58] H: 0.73 I: 0.62	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ No characteristics for the subsample of people with mild to severe dementia reported</li> <li>▪ Data of n = 2 people with dementia were largely incomplete and therefore excluded</li> <li>▪ Mokken analysis was carried out</li> <li>▪ Sample for testing the assessment for very severe dementia included also mild to severe dementia</li> </ul>
				IRR	A: 0.79, 0.88 <sup>h</sup> [0.69, 0.88 <sup>h</sup> ] B: 0.59, 0.74 [0.55, 0.71] C: 0.71, 0.83 [0.68, 0.81] D: 0.68, 0.81	-	<ul style="list-style-type: none"> <li>▪ No sample size calculation</li> <li>▪ No characteristics for the IRR sample of people with dementia and proxy-raters reported</li> <li>▪ Only total ICC and average ICC values without CI were reported</li> <li>▪ No information about blinding reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
	Original version (NL)				[0.68, 0.81] E: 0.47, 0.64 F: 0.72, 0.84 [0.58, 0.73] G: 0.72, 0.84 [0.73, 0.85] H: 0.59, 0.74 I: 0.62, 0.77		<ul style="list-style-type: none"> <li>Data of n = 2 people with dementia were largely incomplete and therefore excluded</li> <li>Sample for testing the assessment for very severe dementia included also mild to severe dementia</li> </ul>
Ettema et al. 2007 [10]	QUALIDEM (NL) <ul style="list-style-type: none"> <li>37 items</li> <li>[18 items for very severe dementia]</li> </ul>		<ul style="list-style-type: none"> <li>Subsample of people with mild to severe dementia (n = 51) and a sample of people with mild to very severe dementia [n = 62]</li> <li>Sample information was equal to the IRR sample</li> </ul>	TRR	A: 0.85 [0.80] B: 0.89 [0.78] C: 0.82 [0.87] D: 0.84 [0.79] E: 0.73 F: 0.85 [0.83] G: 0.74 [0.75] H: 0.88 I: 0.80	?	<ul style="list-style-type: none"> <li>No sample size calculation</li> <li>No characteristics for the TRR sample of people with dementia and proxy-raters reported</li> <li>Only total ICC values without CI were reported</li> <li>No information about blinding reported</li> <li>Data of n = 2 people with dementia were largely incomplete and therefore excluded</li> <li>Second observation after one week</li> <li>Sample for testing the assessment for very severe dementia included also mild to severe dementia</li> </ul>
Bouman et al. 2011 [95]	QUALIDEM (NL) <ul style="list-style-type: none"> <li>37 items</li> <li>[18 items for very severe dementia]</li> <li>Original subscales</li> </ul>	Evaluation of the scalability and reliability of the QUALIDEM	<ul style="list-style-type: none"> <li>Convenience sample based on 4 primary cross-sectional studies in nursing homes resulting in a sample of n = 759 people with mild to severe dementia and n = 214 of people with very severe dementia</li> <li>Inclusion (participants): Dementia diagnosis</li> <li>Exclusion (participants): Diagnosis of Parkinson's disease, a diminution of consciousness or terminal care judged by the nursing home physician (valid for 1 of the 4 primary studies)</li> <li>Inclusion (proxies): Not reported</li> <li>Exclusion (proxies): Not reported</li> <li>Proxy-training: CNAs received 1 hour's oral instruction at the start of the observation period, a written instruction was available (valid for one primary study), two certified nursing assistants rated the QoI independently, discrepancies were discussed until consensus was reached (valid for three primary studies)</li> </ul>	IC	A: 0.86 [0.80] B: 0.90 [0.82] C: 0.75 [0.46] D: 0.73 [0.68] E: 0.66 F: 0.77 [0.40] G: 0.55 [0.57] H: 0.76 I: 0.55	+	<ul style="list-style-type: none"> <li>No characteristics for the proxy-raters reported</li> <li>Mokken analysis was carried out</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Dichter et al. 2011 [94]	<p>QUALIDEM (DE)</p> <ul style="list-style-type: none"> <li>▪ 37 items</li> <li>▪ [18 items for very severe dementia]</li> </ul> <p>Factor analysis:</p> <p>A: Satisfied behaviour (8), [7]</p> <p>B: Unapproachable and satisfied behaviour (9), [4]</p> <p>C: Positive self-image (4)</p> <p>D: Negative affect (3), [4]</p> <p>E: Social relations (4)</p> <p>F: Feeling at home (4)</p> <p>G: Restless tense behaviour (3), [3]</p> <p>H: Having something to do (2)</p>	Investigation of the validity and reliability of the German version of the QUALIDEM	<ul style="list-style-type: none"> <li>▪ Convenience sample based on 2 primary studies in nursing homes resulting in a sample of n = 203 people with mild to moderate dementia and n = 283 of people with severe and very severe dementia</li> <li>▪ Inclusion (participants): MMSE <math>\leq</math> 24, residence duration <math>\geq</math> 2 weeks</li> <li>▪ Exclusion (participants): Diagnosis of schizophrenia or other psychotic disorders (valid for 1 primary study)</li> <li>▪ Inclusion (proxies): Nurses (registered nurses or nursing assistants) must have a close relationship to the particular persons with dementia</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Nurses were guided through each proxy-rating by a trained study assistant</li> </ul>	IC	<p>A: 0.87 [0.83]</p> <p>B: 0.86 [0.72]</p> <p>C: 0.69</p> <p>D: 0.71 [0.62]</p> <p>E: 0.74</p> <p>F: 0.64</p> <p>G: 0.67 [0.61]</p> <p>H: 0.74</p>	?	<ul style="list-style-type: none"> <li>▪ Small sample size for the 37-item version of the QUALIDEM</li> <li>▪ Subscales were based on an explanatory factor analysis</li> <li>▪ No characteristics for the proxy-raters reported</li> </ul>
Dichter et al. 2013 [96]	<p>QUALIDEM (DE)</p> <ul style="list-style-type: none"> <li>▪ 37 items</li> <li>▪ [18 items for very severe dementia]</li> <li>▪ Original subscales</li> </ul>	Evaluation of scalability and internal consistency of the German version of the QUALIDEM	<ul style="list-style-type: none"> <li>▪ Convenience sample based on 3 primary studies in nursing homes resulting in a sample of n = 378 people with mild to severe dementia and n = 256 of people with very severe dementia</li> <li>▪ Inclusion (participants): MMSE <math>\leq</math> 24 or FAST score <math>\geq</math> 2, residence duration <math>\geq</math> 2 weeks</li> <li>▪ Exclusion (participants): Diagnosis of schizophrenia or other psychotic disorders (valid for 1 primary study)</li> <li>▪ Inclusion (proxies): Nurses (registered nurses or nursing assistants) must have a close relationship to the particular persons with dementia</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Nurses were guided through each proxy-rating by a trained study assistant</li> </ul>	IC	<p>A: 0.81<sup>1</sup> [0.73]</p> <p>B: 0.91 [0.86]</p> <p>C: 0.73 [0.50]</p> <p>D: 0.69 [0.59]</p> <p>E: 0.67</p> <p>F: 0.77 [0.53]</p> <p>G: 0.53 [0.42]</p> <p>H: 0.62</p> <p>I: 0.23</p>	+	<ul style="list-style-type: none"> <li>▪ Subscales were confirmed by an confirmatory Mokken scale analysis</li> <li>▪ No characteristics for the proxy-raters reported</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
Graske et al. 2014 [81]	QUALIDEM (DE) ▪ 37 items ▪ [18 items for very severe dementia] ▪ Original subscales	Evaluation of psychometric properties of 3 QoI instruments in German shared housing arrangements	<ul style="list-style-type: none"> <li>▪ Convenience sample recruited in 36 shared housing arrangements in Berlin resulting in a sample of n = 104 people with cognitive impairment or dementia</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Registered nurses working in the recruited shared housing arrangements</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Nurses received a training session prior to data collection including carrying out interviews with people with cognitive impairment</li> </ul>	IC	A: 0.76 [0.71] B: 0.86 [0.79] C: 0.66 [0.47] D: 0.62 [0.67] E: 0.57 F: 0.60 [0.32] G: 0.21 [0.66] H: 0.57 I: 0.18	?	<ul style="list-style-type: none"> <li>▪ Small sample size</li> <li>▪ Subscales were confirmed by a confirmatory factor analysis</li> </ul>
Dichter et al. 2014 [16]	QUALIDEM (DE) ▪ 37 items ▪ [18 items for very severe dementia] ▪ Original subscales	Evaluation of the inter-rater and test-retest reliability of the QUALIDEM	<ul style="list-style-type: none"> <li>▪ Convenience sample recruited in 9 nursing homes resulting in a sample of n = 108 with mild to severe and n = 53 people with very severe dementia</li> <li>▪ QoI of each person with dementia was assessed four times by different proxy-raters</li> <li>▪ Inclusion (participants): FAST score <math>\geq 2</math>, residence duration <math>\geq 2</math> weeks</li> <li>▪ Exclusion (participants): Diagnosis of Parkinson's disease</li> <li>▪ Inclusion (proxies): Nurses (registered nurses or nursing assistants) must have a close relationship to the particular persons with dementia, working contract of at least half time work, nurses had to be at work on most days within the last two weeks prior to data collection</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Nurses were guided through each proxy-rating by a trained study assistant</li> <li>▪ Sample was equal to the IRR sample with the exception of 1 person with mild to severe and 1 with very severe dementia who died between the first and second QoI rating</li> </ul>	IRR	A: 0.61 (0.51-0.69) [0.67 (0.56-0.78)] B: 0.47 (0.37-0.57) [0.28 (0.14-0.44)] C: 0.49 (0.39-0.59) [0.50 (0.37-0.64)] D: 0.60 (0.51-0.69) [0.42 (0.28-0.57)] E: 0.35 (0.24-0.45) F: 0.59 (0.50-0.68) [0.39 (0.26-0.54)] G: 0.55 (0.46-0.64) [0.64 (0.51-0.75)] H: 0.40 CI: 0.29-0.50 I: 0.52 CI: 0.43-0.62	-	<ul style="list-style-type: none"> <li>▪ Sample size was based on a sample size calculation</li> </ul>
				TRR	A: 0.85 (0.79-0.90) [0.70 (0.53-0.82)] B: 0.81 (0.73-0.87) [0.79 (0.66-0.87)] C: 0.75 (0.66-0.83) [0.79 (0.66-0.87)] D: 0.80 (0.71-0.86) [0.78 (0.65-0.87)] E: 0.77 (0.67-0.83) F: 0.84 (0.78-0.89) [0.71 (0.54-0.82)]	+	<ul style="list-style-type: none"> <li>▪ Sample size was based on a sample size calculation</li> <li>▪ Second QoI rating after 1 week</li> </ul>

Source	QoI-Measurement Subscales (No of items)	Objective	Sample (in- and exclusion criteria), Setting Proxy Rating (in- and exclusion criteria, training)	Test	Results $\alpha$ /ICC (95% CI)	Method Rating	Methodological comments
					G: 0.75 (0.65-0.82) [0.74 (0.59-0.84)] H: 0.82 (0.74-0.87) I: 0.88 (0.83-0.92)		
<b>Vienna List</b>							
Porzolt et al. 2004 [44]	Vienna List (AT) ▪ 40 items  A: Communication (15) B: Negative affect (10) C: Bodily contact (5) D: Aggression (4) E: Mobility (6)  Original version (AT)	Development of a new QoI measurement for people with very severe dementia	<ul style="list-style-type: none"> <li>▪ Consecutive sample of n = 217 people with dementia recruited from a Department of Palliative Geriatrics in Vienna</li> <li>▪ n = 771 (total) QoI ratings were performed of the n = 217 people with dementia, n = 386 from nurses and n = 385 from physicians</li> <li>▪ Inclusion (participants): Not reported</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ Inclusion (proxies): Nurses or physicians working in the recruiting Department</li> <li>▪ Exclusion (proxies): Not reported</li> <li>▪ Proxy-training: Not reported</li> </ul>	IC	Nurses: A: 0.93 B: 0.88 C: 0.90 D: 0.86 E: 0.81  Physicians: A: 0.94 B: 0.89 C: 0.90 D: 0.87 E: 0.82	?	<ul style="list-style-type: none"> <li>▪ No information about proxy-rater characteristics, in- or exclusion criteria and training reported</li> <li>▪ No information about the number of repeated QoI ratings per participant reported</li> </ul>
			<ul style="list-style-type: none"> <li>▪ IRR sample of n = 22 people with dementia from the IC sample</li> <li>▪ Inclusion (participants): QoI rating by one nurse and one physician at the same time confirmed by an electronic recording system</li> <li>▪ Exclusion (participants): Not reported</li> <li>▪ For information about proxy-raters see IC above</li> </ul>	IRR	A: 0.71 <sup>d</sup> B: 0.57 C: 0.53 D: 0.35 E: 0.81	-	<ul style="list-style-type: none"> <li>▪ No characteristics of people with dementia and proxy-observers for the IRR sample reported</li> <li>▪ No sample size calculation</li> <li>▪ Only total Spearman rank correlation coefficients without CI were reported</li> <li>▪ No information about blinding reported</li> </ul>

$\alpha$  = Cronbach's alpha, ICC = Intra-Class Correlation coefficient, CI = 95% Confidence Interval, IC = Internal Consistency, TRR = Test-Retest Reliability, DSM IV Criteria = Diagnostic and Statistical Manual of Mental Disorders IV Edition, MMSE = Mini Mental Status Examination, NINCDS-ADRDA = National Institute of Neurologic and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association criteria, ICD-10 = International Classification of Diseases 10<sup>th</sup> Revision, GDS = Global Deterioration Scale, CDR = Clinical dementia rating scale, ICD-9 = International Classification of Diseases 9<sup>th</sup> Revision, DSM II Criteria = Diagnostic and Statistical Manual of Mental Disorders II Edition.

<sup>a</sup>Pearson product-moment correlation coefficients, <sup>b</sup>F-Test statistics (p-value), <sup>c</sup>Range of Cronbach's-Alpha values, <sup>d</sup>Spearman rank correlation coefficient, <sup>e</sup>Cohen's kappa, <sup>f</sup>Cohen's weighted kappa, <sup>g</sup>SEM = Standard Error of the Mean, <sup>h</sup>Average Intra-Class Correlation coefficient, <sup>i</sup>Coefficient Rho.