

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/263395183>

Measuring the quality of life and well-being of people with dementia: A review of observational measures

Article in *Dementia* · June 2014

DOI: 10.1177/1471301214540163 · Source: PubMed

CITATIONS

28

READS

2,383

3 authors:



Katherine Algar-Skaife

Bangor University

7 PUBLICATIONS 39 CITATIONS

[SEE PROFILE](#)



Bob Woods

Bangor University

329 PUBLICATIONS 12,107 CITATIONS

[SEE PROFILE](#)



Gill Windle

Bangor University

77 PUBLICATIONS 2,348 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Project

Created Out of Mind: Shaping perceptions of the dementias [View project](#)



Project

RESPECT trial [View project](#)

Measuring the quality of life and well-being of people with dementia: A review of observational measures

Katherine Algar, Robert T Woods and Gill Windle
Dementia published online 24 June 2014
DOI: 10.1177/1471301214540163

The online version of this article can be found at:
<http://dem.sagepub.com/content/early/2014/06/24/1471301214540163>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Dementia* can be found at:

Email Alerts: <http://dem.sagepub.com/cgi/alerts>

Subscriptions: <http://dem.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://dem.sagepub.com/content/early/2014/06/24/1471301214540163.refs.html>

>> [OnlineFirst Version of Record](#) - Jun 24, 2014

[What is This?](#)

Measuring the quality of life and well-being of people with dementia: A review of observational measures

Dementia

0(0) 1–26

© The Author(s) 2014

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1471301214540163

dem.sagepub.com



Katherine Algar, Robert T Woods and Gill Windle

Dementia Services Development Centre (DSDC), Bangor University, Wales, UK

Abstract

The dynamic nature of psychosocial interventions implies that trying to measure their effects using standardised clinical trial measures may not capture their full effects. Rich and valuable data during the sessions may be missed by using standard quality of life questionnaires. This paper compares observational measures in the context of recording the well-being of a person with dementia during and outside of a visual arts intervention. A literature search was conducted using systematic principles of searching, screening and retrieval to identify peer-reviewed English language evaluations of research projects using observational measures with people with dementia. Psychometric properties, strengths and weaknesses of 11 observational tools are reviewed in order to identify the most appropriate one for evaluating a visual art intervention for people with dementia. This review supports the Greater Cincinnati Chapter Well-Being Observation Tool as an appropriate measure to evaluate a visual art programme for people with dementia. The results of this review will help researchers plan projects to show the full range of effects for people with dementia for taking part in art sessions.

Keywords

dementia, observation, well-being, intervention, review

Introduction

Estimates anticipate a doubling in the number of people with a dementia in the next generation (Alzheimer's Disease International, 2009). As there are no curative treatments as yet for the dementias, it is important to promote well-being and quality of life (QoL) to improve the lives of those living with the condition now. In recognition of this, there has been an increase over the last decade of studies where QoL is viewed as the primary outcome

Corresponding author:

Katherine Algar, DSDC Wales, Bangor University, Arduwy, Normal Site, Bangor LL57 2PZ, UK.

Email: k.algar@bangor.ac.uk

(Ettema, Dröes, de Lange, Mellenbergh, & Ribbe, 2005). The World Health Organisation (WHO) defines QoL as

The product of the interplay between social, health, economic and environmental conditions which affect human and social development. It is a broad-ranging concept, incorporating a person's physical health, psychological state, level of independence, social relationships, personal beliefs and relationship to salient features in the environment. As people age, their quality of life is largely determined by their ability to access needed resources and maintain autonomy, independence, and social relationships. (World Health Organization, 2004, p.48)

This definition highlights that QoL is a multifaceted construct. Non-pharmacological interventions can increase QoL of people with dementia (Olazarán et al., 2010) but measuring such a multi-factorial construct presents many challenges. A number of measures have been developed which measure QoL. Some are generic, meaning that the same measure can be used to measure QoL in people with a variety of different conditions as well as in healthy individuals. For example, the Health Utilities Index (HUI[®]; Feeny, Furlong, Boyle, & Torrance, 1995) which gives a measure of both health status and health-related QoL. Other QoL measures are disease-specific and are designed to be relevant to a certain condition. A Dutch review of QoL measures found six dementia specific QoL measures, three dementia-specific measures related to QoL and nine generic QoL instruments used in dementia research since 1990 (Ettema et al., 2005). They conclude, however, that the generic QoL measures would only be useful when comparing health-related QoL in different populations, and that the dementia-specific measures should be preferred when the research concerns people with dementia primarily.

Of these six dementia-specific measures, one was observational, (Dementia Care Mapping (DCM); Kitwood & Bredin, 1992), one was through self-report, Dementia Quality of Life instrument (DQoL; Brod, Stewart, Sands, & Walton, 1999), two were through a proxy report from the caregiver, Alzheimer's Disease Related Quality of Life (ADRQL; Rabins, Kasper, Kleinman, Black, & Patrick, 1999) and Quality of Life for Dementia (QOL-D; Terada et al., 2002), and two were from an interview with both the patient and caregiver, Quality of Life in Alzheimer's Disease (QOL-AD; Logsdon, Gibbons, McCurry, & Teri, 1999) and The Cornell-Brown Scale for Quality of Life in Dementia (Ready, Ott, Grace, & Fernandez, 2002). All six measures provide quantitative data.

A more recent systematic review searched for non-pharmacological interventions to improve QoL of people with dementia and found 20 relevant studies (Cooper et al., 2012). All but one used validated QoL measures, including the QoL-AD, ADRQL and D-QOL, as well as the Quality of Life Assessment, (WHOQOL: World Health Organization, 1995), health-related quality of life for people with dementia – DEMQOL (Banerjee, Smith, Lamping, Foley, Smith, & Murray, 2004) and the Health Utilities Index Mark 3 (Feeny et al., 2002). All of these measures use self-reported or a proxy report of QoL. Three studies did not use these measures, but instead used observer-rated measures such as DCM and the Quality of Life in Late-Stage Dementia Scale (QUALID; Weiner et al., 2000). Although the QUALID is classified as observer rated by the authors, it seems more appropriately classified as proxy rated QoL as it consists of an interview with an informant about the last 7 days of the person with dementia.

Given the growing number of QoL measures for people with dementia, researchers and clinicians may encounter difficulties in selecting which one might best capture outcomes of interest. To address this, in 2008, a review of outcome measures for psychosocial

interventions was undertaken by INTERDEM, early detection and timely INTERvention for DEMentia; a pan-European research group on early, timely and quality psychosocial interventions in dementia (Moniz-Cook et al., 2008). They suggest that the QoL-AD is the measure of choice for evaluating QoL in people with dementia, as it is short, has been shown to be sensitive to change in an intervention study, can be used internationally as it has been translated into many different languages, and has been shown to correlate with health utility measures.

Most of these self-report and proxy report QoL measures already mentioned have been validated, and shown to be sufficiently sensitive to change in order to evaluate interventions (Cooper et al., 2012); however, there are methodological issues worth consideration. Although it is increasingly recognised that people at all stages of dementia can give their own views on what affects their quality of life (Alzheimer's Society, 2010; Harris, 2002; Keady, 1996; Sabat, 2001; Sterin, 2002), using a self-reported measure could limit the group of people able to be evaluated as a number of measures use exclusion criteria according to cognitive or communication abilities. Therefore, some self-reported QoL measures are not viable for use beyond a certain severity of dementia. If measuring over several time points, disease progression and changes in cognitive function could result in missing data (Fossey, Lee, & Ballard, 2002). It has also been shown that caregiver's proxy reports underestimate ratings of QoL, especially when the participant is more cognitively impaired, and has a poor affect state (Magaziner, Simonsick, Kasher, & Hebel, 1988).

As an alternative, observation is a way of uniquely representing the experience of the person with dementia within the intervention and quantifying it for statistical analysis. Observation presents opportunities to examine real processes and outcomes of specific research interest (Aspland & Gardner, 2003), regardless of cognitive abilities. This is not to say, however, that people with dementia cannot give an account of their own feelings of quality of life, just that perhaps they might struggle with a formal self-report measure.

Clinical anti-dementia drug trials are often derived from a theoretical framework and have a clear aim of improving cognitive abilities, and thus this has typically been the primary outcome (Moniz-Cook et al., 2008), although changes in functional ability are now increasingly required by regulatory authorities. Psychosocial interventions, however, are more complex and often come from theories from different disciplines and constructs, with several aspects to be measured. The dynamic nature of psychosocial interventions implies that trying to measure their effects using standardised quantitative self-report clinical trial measures and approaches may not capture their full effect. Findings such as 'you do it for the moment' (MacPherson, Bird, Anderson, Davis, & Blair, 2009), an evaluation of the National Gallery of Australia's art viewing programme for people with dementia (PwD), which used systematic observation as one of the methodological approaches, suggest that rich and valuable data during the sessions may be missed by using standard QoL questionnaires, especially with more severe levels of dementia. Fifteen participants with dementia attended the gallery access programme from the community and eight from residential care. Observations during the programme showed that while in the gallery, levels of engagement, animation and confidence increased, and that participants engaged in discussions about the artwork. The authors report that these effects did not last, so perhaps these positive effects would have been missed if the researchers had chosen to use one of the standardised QoL measures already discussed.

There is a growing interest in the use of art, such as the National Gallery of Australia programme, as a psychosocial intervention for people with dementia. Care providers are

increasingly looking for novel ways to improve health and well-being, and deliver quality services across the community. In doing this, they have begun to consider innovative approaches such as involvement in the creative arts, as the recent report 'Creative Homes' (Baring Foundation, 2011) suggests.

In a review of creative therapies for people with dementia (including music, art, drama and dance therapy) 13 studies were identified (Salisbury, Algar, & Windle, 2011). Many positive effects were found from participation, such as the improvement of interaction skills and people coming to terms and coping better with dementia. However, overall, the studies were of poor quality in that the art interventions needed further clarification and methods of evaluation needed development and validation. This has been further supported by a systematic review of art therapies and dementia (Beard, 2011) which concludes that the existing evidence lacks adequate study design, with measurement tools (if any) unspecified, emphasising clinical outcomes rather than investigating quality of life, and lacks adequate analysis of the data.

Therefore, to add to the evidence base, research examining the creative arts with people with dementia should measure more than clinical outcomes and look at QoL. In addition, rather than rely on self-report of QoL, observational methods could capture the unique effects the arts might have on the person, such as increasing engagement, activity and social interaction. This may provide additional, rich data beyond self-reported outcome measures. Observation therefore offers another way of evaluating psychosocial interventions rather than QoL questionnaires pre- and post-intervention. It gives an opportunity for evaluation of QoL within the intervention, as well as pre and post if required.

In dementia care research, observational measures have been used for some time to assess the quality of care the person is receiving. However, as demonstrated by MacPherson et al. (2009), systematic observation also has the capacity to be applied in other settings. An earlier review of observational studies (Brooker, 1995) identifies a number of approaches that could potentially be applied in an intervention study. This early review focussed on observing the quality of institutional care and highlights important behaviours to capture, such as engagement, activity and social interaction. It also introduces DCM, which at the time was a new approach to capturing a wide range of behavioural categories based on Kitwood's social psychological theory of person-hood in dementia.

Despite this early contribution, researchers, practitioners or clinicians often have limited literature to consult and inform their choice of the most appropriate observational method and measure that could be applied in a psychosocial intervention setting, especially one involving creative activities. This paper aims to address this need for information. It will review observational approaches that could be applied to psychosocial intervention studies involving creative activities. This will update and augment the evidence provided in the early review by Brooker (1995). Specifically it aims to:

- (1) Identify and compare the strengths, weaknesses and appropriateness of observational measures for recording the well-being of a person with dementia.
- (2) Develop research and practice implications and recommendations.

Method

A literature search was conducted using systematic principles of searching, screening and retrieval to identify peer-reviewed English language evaluations of intervention studies using

observational measures with people with dementia. The academic databases PsycINFO, Web of Science, CINAHL and ASSIA were searched on 7 July 2012 and again on 22 May 2013. Combinations of the following terms were used as thesaurus terms or keywords: 'dementia', 'Alzheimer's', 'well being', 'quality of life', 'intervention' and 'observation'. Reference lists were scanned for relevant papers. Discussions were also held with experts in observational methods, websites of known researchers/research centres searched and where necessary, email correspondence with researchers exchanged. A search of internet search engines was also conducted. Papers were excluded if they were not in the English language, were not relevant, or were not using observational measures in a research capacity for measuring well-being or quality of life (i.e. were about pain assessment).

Results

A total of 2574 papers were identified using the search terms described above. Removing duplicates and screening titles and abstracts left 43 potential papers. Full text was obtained and a further 15 papers rejected as they did not meet the inclusion criteria. This left 28 eligible papers. In total, 11 observational measures were included for appraisal and seventeen supporting papers included.

Table 1 gives an overview and summary of the 11 observational measures found in the search. The authors were unable to obtain the original development papers for two of the measures (INTERACT and QUIS); therefore, the supporting papers are used for review purposes. Table 2 gives a summary of the psychometric properties of each measure in terms of inter-rater reliability, test-retest reliability and validity. The following also discusses other important factors to consider when choosing a measure, such as training requirements, sensitivity to change and practical considerations.

Inter-rater reliability

Inter-rater reliability can be defined as the reproducibility of the scores between multiple raters (Kline, 1979). Cohen's kappa is most commonly used to report inter-rater reliability and refers to the proportion of responses that both raters agree (Streiner & Norman, 2003). Altman (1991) suggests the following guidelines by which the value of kappa indicates the strength of agreement (p. 404): Poor <0.20; Fair, 0.21–0.40; Moderate, 0.41–0.60; Good, 0.61–0.80; Very good, 0.81–1.00. All but one of the identified measures (Bradford Well-Being Profile) reports inter-rater reliability statistics. Five of the remaining 10 measures (AwareCare, DCM, Greater Cincinnati Chapter Well-Being Observation Tool, OERS, & QUIS) report Cohen's kappa coefficient, ranging from 0.54 to 0.89, which according to Altman, indicates that they all show moderate to very good inter-rater reliability.

As shown in Table 2, the highest inter-rater reliability was demonstrated in the OERS, which indicated good to very good agreement.

The QUIS and AwareCare are both reported to have a Cohen's kappa of 0.75 or above, showing good inter-rater reliability. In AwareCare, inter-rater reliability was calculated by both researchers observing 12 participants, three from each participating home, at the same time on five occasions (Clare et al., 2012).

In the Greater Cincinnati Chapter Well-Being Observation Tool, inter-rater reliability was measured by having two trainers observe the same five participants during three Memories in the Making[®] sessions and three other activities sessions. Despite the fact,

Table 1. An overview of observational measures for people with dementia.

Measure	Purpose of the measure	Domains observed	Observation procedure	Key references
AwareCare	To aid identification of behavioural signs of awareness in people with severe dementia	<p>Events that happen:</p> <ul style="list-style-type: none"> Someone is nearby; Resident is touched; Resident is spoken to; Talking nearby; Loud noise; Object nearby; Food/drink <p>Introduced events: Call by name; Take hand; Introduce one object (picture, lavender pillow, or textured cushion); Or introduce a more personal object</p> <p>Response: Eyes flicker; makes eye contact; explores with eyes; smiles; frowns; nods or shakes head; moves head; reaches; grasps or holds; moves towards; moves away; single words; mumbling; shouts or moans</p>	Five separate 30-minute observations made in communal areas while resident is awake, at different times of day. A record is made of each stimulus that happens in the observation session, and all responses, as well as a note about the setting, the resident's behaviour at the beginning of observation, any changes or events that happen during observation and any other behaviours not covered in the matrix	Clare et al. (2012)
Behaviour observation	Developed initially to code behaviour from video of people with dementia during an art gallery visit, and later used to evaluate Snoezelen room vs. garden	<p>Very engaged, engaged, neutral, disturbed/disengaged</p>	Time sampling for 2 minutes of 5 seconds 'on' (watching the participant), 5 seconds 'off' (coding previous 5 seconds)	MacPherson et al. (2009)

(continued)

Table 1. Continued.

Measure	Purpose of the measure	Domains observed	Observation procedure	Key references
Bradford well-being Profile	To provide a simpler tool than DCM to indicate the well-being of a care home resident	Can communicate wants, needs & choices; makes contact with other people; shows warmth or affection; shows pleasure or enjoyment; alertness, responsiveness; uses remaining abilities; expresses self creativity; is cooperative or helpful; responds appropriately to people/situations; expresses appropriate emotions; relaxed posture or body language; sense of humour; sense of purpose; signs of self respect	Each indicator scored as 0 – no sign, 1 – some signs, 2 – significant signs	Bruce (2000) Bradford Dementia Group (2008)
Creative-Expressive Abilities Assessment tool (CEAA)	To record 'in-depth' information about the creative abilities of people with dementia.	27 Items related to memory, attention, language, psychosocial, reasoning & problem solving, emotion and culture	Each item is assigned a code to indicate the frequency the behaviour is shown during the creative activity: 0 = not observed; 1 = Never; 2 = Rarely; 3 = Sometimes; 4 = (nearly) Always Up to four participants can be observed at one time and recorded on one score sheet.	Gottlieb-Tanaka et al. (2008a). Gottlieb-Tanaka et al., (2008b) http://www.dementia-activities.com/CEAA/product_dementia_assessment_tool.html

(continued)

Table 1. Continued.

Measure	Purpose of the measure	Domains observed	Observation procedure	Key references
Dementia Care Mapping (DCM)	To record the quality of care and quality of life for the person with dementia living in a care setting	(DCM 8) Behavioural Category Codes (BCC – Type of behaviour or activity engaged in): Articulation; Borderline; Cool; Doing for self; Expressive; Food; Going back; Intellectual; Joints; Kum and Go; Leisure; Nod Land Of; Objects; Physical; Religion; Sexual expression; Timalation; Unresponded to; Vocational; Withstanding; Excretion; Yourself; Zero option Mood/Engagement Code (ME – judgement of affect and engagement): +5 Very positive + 3 Considerable signs of positive mood + 1 Neutral – 1 Small signs of negative mood – 3 Considerable signs of negative mood – 5 Very negative	Observation of five to eight people continuously for six hours in a communal area. After every 5-minute time frame, the observer records two codes – BCC and ME for each participant. The mapper also records when Personal Detractions (PD) or Personal Enhancers (PE) occurs.	Kitwood & Bredin (1992)

(continued)

Table 1. Continued.

Measure	Purpose of the measure	Domains observed	Observation procedure	Key references
Greater Cincinnati Well-being Observation Tool	To measure seven domains of well-being in people with dementia	Indicators developed for each domain of well-being as derived from Lawton's (1991) conceptualisation: Interest; sustained attention; pleasure; negative affect; sadness; self-esteem and normalcy added	Observations for up to three participants recorded every ten minutes – each indicator scored using scale ranging from 0 (never demonstrates) to 4 (always demonstrates)	Rentz (2002) Kinney and Rentz (2005)
INTERACT	To measure the effects of Snoezelen on people with dementia	22 items relating to mood, speech, relating to other people, relating to the environment, need for prompting, and stimulation level.	An observer watches video of Snoezelen session and scores each item using 5-point Likert scale ranging from 'not at all' to 'nearly all the time'	Baker and Dowling (1995)
Menorah Park Engagement Scale (MPES)	To quantify the varying levels of engagement during activities	Constructive engagement (CE; any verbal or motor behaviour in response to activity); Passive engagement (PE; listening and/or looking behaviour in response to activity); Non-engagement (NE; staring into space, sleeping or motor/verbal behaviour in response to activity not currently participating in); Self-engagement (SE; motor/verbal listening and/or looking	Coding length of time in each type of engagement was exhibited during 10-minute observation windows: 0 = never seen; 1 = seen up to half of the activity time; 2 = seen over half the activity time.	Judge et al. (2000)

(continued)

Table 1. Continued.

Measure	Purpose of the measure	Domains observed	Observation procedure	Key references
Observed Emotion Rating Scale (OERS)	To measure the affect of a person with dementia by observing facial expression, tone of voice, and body movements	behaviour while activity not on or when choosing not to participate) Pleasure; general alertness; anger; anxiety or fear; sadness	After a 10-minute observation, the rater chooses one of six possible time interval codes (7 = not in view, 1 = never, 2 = less than 16 seconds, 3 = 16–59 seconds, 4 = 1–5 minutes, 5 = more than 5 minutes)	Lawton et al. (1996) Lawton et al. (1999)
Positive Response Schedule for Severe Dementia (PRS)	To evaluate an intervention for people with severe dementia	Engagement: deliberate body movement, deliberate head movement, vocalisation, looks at environment, looks at carer; initiates interaction, engagement Emotion: happy, sad, fear	Suggested continuous observations 20 minutes baseline, 20-minute intervention, 20-minute postintervention. Rated at 20-second intervals, with 10 seconds to record. Only the first instance of behaviour in any one time frame is recorded.	Perrin (1997)
Quality of Interaction Scale (QUIS)	To assess the number and quality of interactions between people with dementia and staff in a residential setting.	Positive social; positive care; neutral; negative protective; negative restrictive	Series of ten 15-minute observations across the working day, over a period of approximately two weeks.	Dean et al. (1993)

Table 2. Summary of the psychometric properties of the observational measures.

	Inter rater reliability	Test–retest reliability	Validity
AwareCare	Inter-rater reliability consistently high (mean > 0.6) for majority of responses Response to spontaneous stimuli – Cohen’s kappa 0.75. Introduced stimuli – complete agreement that given stimulus occurred.	Data were examined but no kappa value reported	Face validity: Focus groups were held with care staff, family members and care home managers Construct validity: Assessed using correlational analysis in terms of the association with dementia severity, cognitive ability and functional behaviour as the theoretical framework of awareness predicts that a greater severity of dementia would mean fewer signs of awareness. Results showed that participants showing more signs of awareness showed less indications of impairment on the other measures. Concurrent validity: Positive correlations between PRS and AwareCare scores indicated that participants giving more positive responses on the PRS also showed more responsiveness on AwareCare.
Behaviour Observation	Kendall’s tau-b = 0.72 ($p < 0.01$)	Not reported	Not reported
Bradford Well-being Profile	Not reported	Not reported	Not reported
CEAA	Kappa values ranged from 0.20 to 0.75. For agreement of total score:	Not reported	Not reported

(continued)

Table 2. Continued.

	Inter rater reliability	Test-retest reliability	Validity
DCM	<p>Cramer's V score = 0.825 Contingency coefficient = 0.975</p> <p>Fossey et al. (2002): A minimum kappa value of <0.8 was established</p> <p>Sloane et al. (2007): An overall agreement for Behavioural Category Codes of 62% with a kappa of 0.54 ($p < 0.01$)</p>	<p>Considerable well-being: $r = 0.58, p < 0.0001$</p> <p>Activities: $r = 0.40, p = 0.003$</p> <p>Social withdrawal: $r = 0.33, p = 0.007$</p>	<p>Concurrent validity: Fossey et al. (2002): Well-ill being score was strongly correlated with QoL derived from Blau scale ($r = 0.73, p < 0.0001$).</p> <p>Sloane et al. (2007): Question the concurrent validity when they found a higher correlation with proxy than direct QoL assessments.</p>
Greater Cincinnati Chapter Well-being Observation Tool	<p>Kappa coefficient could not be generated for 50% of observations because of perfect agreement or only one discrepancy. Kappa coefficient for the remaining observations was 0.65</p>	<p>Not reported but discussion of why not attempted.</p>	<p>Not reported</p>
INTERACT MPES	<p>Mean Pearson's $r = 0.83$</p> <p>90% agreement over 25 observation windows</p>	<p>Not reported</p> <p>Not reported</p>	<p>Not reported</p> <p>Face validity: Extensive discussions were held with activities and nursing staff along with long observations of residents to choose the categories.</p>
OERS	<p>Kappa coefficient ranging from 0.76 to 0.89 in all six affect states</p>	<p>Not reported</p>	<p>Concurrent validity: Shown by comparing scores with ratings of QoL by staff and family members.</p>
PRS	<p>Mean of 80% agreement</p>	<p>Not reported</p>	<p>Face validity: Recognised by how closely the measure is linked to the phenomena being investigated</p> <p>Construct validity: Said to be shown by the</p>

(continued)

Table 2. Continued.

	Inter rater reliability	Test–retest reliability	Validity
			theories that the measure is rooted in engagement theory in 7 items, and theory of non-verbal expression of emotion in three items.
			Concurrent validity: Acknowledged but lack of measures to compare with
QUIS	Observation and coding consistently produce Kappa coefficients of above 0.75	Not reported	Concurrent validity: Ratings of increased quality and quantity of interactions were associated with ‘improvements in ratings of residents’ cognitive impairment, observed depression and functional capacities’ (p. 824).

a kappa coefficient could not be generated for 50% of the observations because of perfect agreement, or only one discrepancy, a good inter-rater reliability was found for the remaining observations (Kinney & Rentz, 2005).

Various inter-rater reliability scores were found for DCM. In a paper establishing the psychometric properties of DCM (Fossey et al., 2002), in Cohort A, a good to very good inter-rater reliability was established between individual raters during pilot mapping sessions. However, in discussing DCM as a research tool, a moderate inter-rater reliability was reported (Sloane et al., 2007). When validating DCM8, Brooker and Surr (2006) report that an inter-rater reliability concordance of 70% was established between the mappers in their study.

Inter-rater reliability is reported in the remaining five measures in a number of ways. The Behaviour Observation in MacPherson et al. (2009) reports acceptable inter-rater reliability as using Kendall’s tau-b. The CEAA reports kappa values ranging from poor to moderate inter-rater reliability but also boasts good agreement on total scores using Cramer’s V score and contingency coefficient (Gottlieb-Tanaka, Lee, & Graf, 2008a). The INTERACT scale reports good inter-rater reliability found when two observers were blinded to the group (control or intervention), and watched the videos twice before scoring (van Weert, van Dulmen, Spreeuwenberg, Ribbe, & Bensing, 2005).

The remaining two measures, the MPES and the PRS, report inter-rater reliability as a percentage. Ninety percent agreement over 25 observation windows was reported in the MPES (Judge, Camp, & Orsulic-Jeras, 2000), and consistent average inter-rater reliability

of 80% was shown in the development of the PRS (Perrin, 1997). However, Hadley, Brown and Smith (1999) achieved inter-rater reliability of at least 99%.

Despite the inconsistency in the way it is reported, all of the measures reporting inter-rater reliability show good agreement.

Test–retest reliability

Test–retest reliability can be defined as the reliability of a measure when used or administered on two occasions separated by a short interval of time. The recommended interval of time is dependent on the type of measure, but between two and 14 days is suggested (Streiner & Norman, 2003).

Test–retest presents challenges when applied to people with dementia, as the behaviour of a person with dementia can vary greatly from hour to hour, day to day. Even if two observations are done at exactly the same time on the same day of the week, a week apart, there are many other factors that could affect behaviour, which presents difficulties in interpreting these statistics. Discussing why they did not attempt to measure test–retest reliability, Kinney and Rentz (2005) note it was due to ‘the inherent variability in the behaviour of individuals with dementia’ (p. 222). As such, test–retest reliability is only reported for two of the measures (AwareCare and DCM).

Clare et al. (2012) examined test–retest reliability during the development of the AwareCare tool, comparing two observations. The fifth observation was carried out at the same time of day, a week later than the fourth observation, so that a direct comparison could be made. Data from all 40 of the participants were examined as to the number of times each stimulus was present at each time point, and the number of times that the stimulus elicited the same response in both sessions. They found considerable variability and do not report a Cohen’s kappa value. Echoing Kinney and Rentz’s (2005) view, the authors comment that the test–retest was affected by the naturally occurring variation across different observations despite being held at the same time of day.

Test–retest reliability is also reported for DCM (Fossey et al., 2002). They measure the test–retest reliability by determining the correlations between the key DCM indices using Spearman’s *r*. They report a good level of test–retest agreement for considerable well-being and a more moderate correlation for Activities and Social withdrawal.

Although considered a good indicator of reliability for quantitative measures, test–retest reliability may be less useful in assessing the quality of an observational measure, especially with people with dementia.

Validity

Validity refers to how accurately a measure or test measures what it actually aims to measure (Kline, 1979). There are several types of validity and some are discussed below. The extent to which the measures are validated varies. The Bradford Well-being Profile is not validated.

Face validity is addressed in three of the measures – AwareCare, MPES and the PRS. This is a measure of whether the test seems to measure what it is intended to. Both AwareCare and MPES address face validity by involving care staff in the development of the measure. In AwareCare, the authors recognise that some of the items with lower kappa values were kept due to the importance placed by staff in the focus groups (Clare et al., 2012).

Construct validity refers to what extent items of a measure represent the construct it claims to measure and relates to other measures in a manner that is consistent with theoretically derived hypotheses (Terwee et al., 2007). It is addressed in two measures – AwareCare and the PRS.

Content validity refers to the extent to which a measure represents all facets of the construct it is designed to cover (Terwee et al., 2007). Although not expressly addressed, content validity is evident in three of the measures (DCM, Greater Cincinnati Chapter Well-Being Observation Tool) due to their strong correlation with theoretical constructs. DCM was grounded in Kitwood's theory of personhood and person-centred care and the Greater Cincinnati Chapter Well-Being Observation Tool was developed using Lawton's (1991) domains of QoL. Lawton's own measure, OERS is rooted in Lawton's (1983) dual-channel hypothesis. This suggests that positive affect is related to engagement in an external event such as a recreational activity, whereas negative affect is related to internal phenomena such as a memory or thought. In developing the OERS, Lawton was trying to test his hypothesis that affect is a result of antecedent events.

Concurrent validity shows the extent to which scores accurately estimate an individual's current state and can be measured by comparing scores with similar measures. This is addressed in five of the measures – AwareCare, DCM, OERS, PRS and QUIS.

During the development of the PRS, Perrin suggests that there is no way to establish concurrent validity due to the lack of measures to compare with as the PRS was developed in recognition that the DCM was not sensitive enough to pick up subtle changes shown in people with severe dementia. The PRS was, however, a good measure to compare with AwareCare when it was developed.

Although not stated as concurrent validity, authors suggest that evidence of validity of the QUIS can be shown by comparison with other outcome measures (Dean, Proudfoot, & Lindsay, 1993).

An inspection of the extent to which the measures are validated demonstrates that AwareCare is the only measure to address face, construct and concurrent validity.

Performance of observation measures in previous intervention studies/sensitivity to change

AwareCare. Although this tool has proved successful in showing changes in behavioural responses by people with severe dementia, it was developed to aid the training of staff to observe responses from residents rather than as a tool for data collection. The authors state it has potential as a research tool, but as yet this has not been tested.

Behaviour Observation. The Behaviour Observation was used when evaluating the effects of taking PwD around the National Gallery of Australia (NGA) to discuss artworks (MacPherson et al., 2009). There were 15 PwD from the community, and eight from residential care in the group. The group sessions were videoed and week 1 and week 5 rated by two independent observers to show changes over time. The two raters coded participant behaviour separately and then arrived at a consensus of definitions for the range of behaviours indicating affect. The results showed a high engagement from the beginning. In week 1, 84% of the observations fell into the engaged or highly engaged category. There was a significant increase in the number of highly engaged observations among the participants in residential care between week 1 and week 5. It was also used live in

real time, as opposed to making observations from a video, to investigate the effectiveness of a Snoezelen room in residential care. Although no significant difference was shown between the Snoezelen room and the other condition (being in the garden), there are many possible reasons given for this, and does not seem to reflect on the sensitivity of the measure (Anderson, Bird, MacPherson, McDonough, & Davis, 2011).

Bradford Well-being Profile. Despite a suggestion that the tool could be used to evaluate the impact of interventions (Bruce, 2000) it was not found in any peer-reviewed research articles.

CEAA. The user guide suggests that the tool is useful for researchers to answer questions such as effectiveness of one creative activity compared to another in increasing expressive abilities and QoL, the most effective intensity of a programme, and whether a programme is suitable for people with different stages of dementia. However, the searches for this paper found no peer-reviewed journal article using the CEAA.

DCM. Although originally developed as a way of assessing the quality of care, DCM has increasingly become a tool used in research, including for the evaluation of an intervention. In a review of published DCM literature, 10 articles were found using DCM to evaluate the impact of different interventions on the person with dementia (Brooker, 2005). Brooker and Duce (2000) use the WIB value to compare the well-being of 25 PwD in three different activities – group reminiscence therapy, structured goal-directed group activity and unstructured time. Researchers used the WIB value as indicators of well-being in each activity. A significant difference was shown between the activities with higher levels of well-being shown in the reminiscence therapy. The authors conclude that DCM was sensitive enough to discriminate between the different activities and suggest that DCM could be used in ‘assessing the impact of therapeutic activity with this client group’ (Brooker & Duce, 2000, p. 358). Brooker (2005) also concludes her review of DCM in research saying that the measure is suited to smaller studies of within-subjects evaluations of comparisons of group interventions, as it has demonstrated the ability to discriminate between various interventions.

The Greater Cincinnati Chapter Well-Being Observation Tool. This tool was developed and piloted to evaluate ‘Memories in the Making[®] (MIM), an arts programme for PwD led by an artist facilitator to encourage self-expressions through the visual arts (Rentz, 2002). A significantly higher proportion was shown in five of the seven well-being domains in the art session compared to the other activity. These domains were interest, sustained attention, pleasure, self-esteem and normalcy. The study showed that the tool was suitable to assess the well-being of PwD, and that it showed differences in well-being within and between the two types of activities in which the participants were observed (Kinney & Rentz, 2005). Another recent study used the Greater Cincinnati Chapter Well-Being Observation Tool to evaluate whether well-being was affected beyond the MIM sessions (Gross, Danilova, Vandehey, & Diekoff, 2013). Seventy-six residents from four long-term care facilities took part in a 12-week MIM programme. Interns rated participants within the sessions in weeks one, six and twelve, and staff from the long term care facility rated participants outside of the session ‘on or about each of the same days as the interns . . . at times that were convenient to them’ (Gross et al., 2013, p. 8). Despite questioning the psychometric properties of the tool itself, they found that participants showed significant improvements over time within the sessions in the same

domains as reported in the previous evaluation by Kinney and Rentz (2005). They did not find any significant changes on any of the domains on ratings made outside of the MIM sessions. However, methodological flaws such as not having the same rater inside and outside of sessions, as well as ratings outside of sessions being made at inconsistent times, could put these results into question.

INTERACT. A study investigating the effect Snoezelen, or Multi Sensory Stimulation (MSS), integrated into 24-hour care, had on mood and behaviour of PwD living in a residential setting used video recordings of morning care and analysed them using the INTERACT scale (van Weert et al., 2005). Results demonstrated that the group receiving the intervention showed

More happiness and enjoyment, related better to the CNA [Certified Nursing Assistants], were more responsive to speaking, and talked more frequently with normal length sentences than the control group. They were also in a better mood and showed less sadness, bored and inactive behaviour, negativism, and reluctance. (van Weert et al., 2005, p.30)

MPES. The MPES was developed to evaluate the effects on engagement during Montessori-based activities for people with dementia in adult day care compared with regular activities (Judge et al., 2000). A significant increase of Constructive Engagement was found during Montessori-based activities compared with regular activities. Low numbers of observations in the domains of Self-Engagement and Non-Engagement meant that they were not formally analysed. However, it was noted that there were no instances of these two categories at all during the Montessori based activities. The MPES was also used in a similar study comparing Montessori based activities with regular activities (Jarrott, Gozali, & Gigliotti, 2008). Again, participants showed more Constructive Engagement in the Montessori based activities, demonstrating the sensitivity of the measure between different activities.

OERS. The OERS is a well-used observational tool. It is one of the outcome measures used in an evaluation of *TimeSlips*, a storytelling programme for PwD (Phillips, Reid-Arndt, & Pak, 2010). The intervention group were observed during the programme and the control group during mealtime. Analysis showed that people in the intervention group had significantly higher pleasure than the control group. However, the authors note that four of the categories, anxiety, sadness, anger and general alertness, were not analysed due to extreme sparseness in distribution as not many people showed these features. This suggests that perhaps the categories are not sensitive enough or too broad.

PRS. Hadley et al. (1999) evaluated individualised interventions for people with severe dementia using the PRS. As the PRS is relatively labour intensive, they undertook two case studies to determine whether the results from the tool could justify using it in a larger scale study. They conclude that the PRS showed more behaviours than DCM. However, they do comment that they found difficulty in differentiating between some of the categories, for example between 'deliberate head movement' and 'looks at environment'.

QUIS. The QUIS was used in an evaluation of two residential units for older people with dementia (Dean, Briggs, & Lindesay, 1993). Observations were made at baseline while the participants were in long stay dementia wards, and at 3, 6 and 12 months once they had

moved to a domus unit. An increase of positive interactions was shown from 49.3% on the long stay ward to 99.2% in the domus unit 12 months later. This demonstrates the tool is sensitive enough to show the changes from poor to good care. However, at such a high level in the domus unit, the tool would no longer be able to show any improvements due to a ceiling effect.

Training requirements

Training requirements vary from measure to measure. For five of the measures (Behaviour Observation, INTERACT, MPES, PRS & QUIS) there is no formal training, but all mention the importance of inter-rater reliability, so it would be assumed that anyone intending to use the measure should practice until they achieve good inter-rater reliability.

A discussion of training is not mentioned in the development of the AwareCare tool, but they do state that the next stage of the project would include training sessions for care staff. The tool is freely available on the internet (<http://reach.bangor.ac.uk/AwareCare.pdf>) with guidelines on how to use it.

No formal training was found for the Bradford Well-being Profile, but the measure comes with guidelines and instructions of how to use it, and it is suggested that training is given to anyone using it.

The CEAA is copyrighted and can be purchased from a website (http://www.dementia-activities.com/CEAA/product_dementia_assessment_tool.html). The tool comes with a user guide and two training DVDs.

To be able to use the DCM, training from a Bradford Dementia Group approved trainer is essential. The Bradford Dementia Group provide an intensive four day course which ends in a formal exam where a pass mark of over 60% is required to be able to go on to use DCM in practice. They also offer a one-day update for those who do not use the tool regularly and need refreshing. Costs and availability can be found on their website.

The Greater Cincinnati Chapter Well-Being Observation Tool can be requested from the second author, Clarissa Rentz. The authors note that although there are detailed operational definitions for each indicator of well-being that were easily interpreted, quite a lot of training was needed for observers to be able to accurately estimate the extent to which participants showed each indicator. In other words, training was needed in deciding between the different categories on the five-point Likert scale (Kinney & Rentz, 2005).

In the development of the OERS, a month of training was provided to the observers, including a discussion of emotions and how they manifest, and watching a video showing facial expressions. Informal rating sessions then took place with a formal test of inter-rater reliability before study data began. Although the scale is now freely available online (from www.abramsoncenter.org/PRI), by the request of Lawton himself, it is suggested those wishing to use the scale watch a training video developed by the Philadelphia Geriatric Center (Lawton et al., 1999).

Practical considerations

There are some practical issues worth considering when deciding which observational measure is most suited for an intervention. In particular, the number of people that can be observed in one sitting will determine how long is needed to capture data for every participant, and also how many raters are needed. Not all of the measures report this. Of

those that do, DCM reports that between five and eight people can be observed at once, the CEAA suggests up to four participants, and the Greater Cincinnati Chapter Well-Being Observation Tool suggests up to three participants are observed at once. This needs to be considered along with the suggested observation window. For example, DCM is continuous observation for six hours, whereas the MPES and OERS are done in 10-minute observations. Another consideration is whether the observations will be done in vivo or from a video. The Behaviour Observation and INTERACT measures both report that the observations were made from a video. This has the advantage that inter-rater reliability can be more easily determined as it can be assumed that raters are watching exactly the same footage at the exact time. However, relying solely on video footage could lead to data being missed due to technical problems, or even simple things such as someone sitting in front of the camera and blocking the view of other participants.

Discussion

This review responds to the apparent gap in resources for researchers wishing to choose a suitable measure to record the well-being of people with dementia during a psychosocial intervention. Eleven observational measures were identified. A review of psychometric properties, as well as sensitivity to change, training requirements and practical considerations has revealed strengths and weaknesses for each measure. Out of all of the measures discussed, the AwareCare tool is the only measure to have reported all aspects of validity and reliability, but has yet to have been used in a peer-reviewed intervention study due to being relatively new.

Two of the measures have been developed to be most suited for use with people with severe dementia. The PRS has been reported to be effective in showing the effects of interventions for people with severe dementia. However, it does seem that a lot of the engagement items depend on the physical abilities of the participant, which could affect results. This leads to the AwareCare tool that has already shown to have good inter-rater reliability and validity. Although it was developed to give care staff a tool to aid the identification of behavioural signs of awareness in people with severe dementia, the authors suggest it could be used as a research tool. Although it is measuring awareness, and not engagement, perhaps the responses could be useful in indicating engagement in people with severe dementia, and it may be useful in an intervention setting.

When considering a psychosocial intervention for people with dementia, these two measures would probably be most suited in a one-to-one setting with people with severe dementia, or if in a group setting, it should only have people with severe dementia. If used in a group setting with participants with mixed abilities, a ceiling effect could be shown for some of the more able participants. Although appropriate for people with severe dementia, neither tool seems to cover as many aspects of QoL for evaluating a visual art programme as some of the other measures.

Of the remaining measures, the QUIS is less suited for evaluating a psychosocial intervention. The QUIS has been adapted by David Sheard (2008) and published by the Alzheimer's Society as an evaluation tool to 'capture through observation the lived experience of PwD living or spending time in a care setting' (p. 2). The time sampling method has been changed to every five minutes for a minimum of 2 hours. On top of this, a 'script' is written to record any quotes or notes of experiences that fall outside of the coding. This version of the QUIS seems to give a less labour intensive alternative to DCM.

Although the QUIS has been shown as a good tool for assessing the care people are receiving in a residential setting, and an alternative to DCM, it seems less suited to assessing an intervention. It could be argued that the interactions recorded could be with the facilitator, rather than with care staff, but work to adapt the scale would be needed to make it more suitable for an intervention, and in turn for any creative activities. For example, the definitions of positive care and negative protective items might need to be changed to relate to the intervention rather than the care they are receiving.

DCM was the method of observation most commonly adopted in the papers identified in the searches. It should however be noted that caution is needed when reporting reliability and validity scores of DCM, as DCM 8 replaced the 7th version in 2005, and the majority of studies reporting these figures are likely to have used DCM 7. However, Brooker and Surr (2006) compared concurrent validity between DCM 7 and DCM8 and a Pearson correlation coefficient of 0.97 ($p < 0.0001$) was calculated between the individual WIB scores on DCM 7 and 8. They concluded that it could be assumed that the validation of DM8 against other measures of quality of life and quality of care would be similar. They state that this would however need to be empirically tested, along with tests of inter-rater reliability, test-retest reliability and internal consistency.

DCM gives a wealth of data and has been shown to be sensitive to changes in QoL within as well as between different activities. It has the advantage that five to eight participants can be observed at a time, compared with other measures where only one participant is observed at a time. However, it is the most time consuming and most expensive measure to train in. A shorter version developed for use in research might make it more suitable in a psychosocial intervention.

The Bradford Well-Being Profile was developed, as it was felt that something was needed that did not require such an extensive input of time and resources as the DCM and that could provide an indication of the well-being of care home residents. However, although psychometric testing of the well-being indicators in a previous version did meet validity criteria, the authors advise caution in interpreting them. They state that 'the tests of validity conventionally used depend on making mathematical assumptions that may well be unfounded when measuring psychological attributes like well-being' (Bradford Dementia Group, 2008, p. 24). They also state that as a psychological attribute, well-being is not something that can be accurately measured meaning issues of inter-rater reliability challenge the validity (Bradford Dementia Group, 2008). For this reason, they removed the scoring system. This means that the newest version of the profile is more of a qualitative tool (Bradford Dementia Group, 2008). The guidelines state that the profile is not suited to be used as an outcome measure for changes in affect during an intervention. Therefore, unless the earlier version was used, this tool is no longer a suitable option for use in a psychosocial intervention.

Behavioural Observation, INTERACT, MPES and OERS have all been used in evaluations of psychosocial interventions previously, suggesting that they are suitable. The Behavioural Observation was used during an art gallery visit for people with dementia, INTERACT in evaluating Snoezelen, the MPES in evaluating Montessori-based activities and the OERS for the evaluation of *TimeSlips*.

The Behavioural Observation was done from video after the event. This provides an easy way to measure inter-rater reliability as each independent rater can be shown exactly the same clip. However, it is more time consuming to rate after the event and relies on the video. Technical issues could mean a loss of data, which may not be detected until after the observational period. However, the same measure was used for live observations to

evaluate Snoezelen rooms and no critique was given to suggest that this caused any problems.

It was noted that the categories in the Behaviour Observation were similar to those of the MPES, although they were conceived independently. The MPES has been shown to be sensitive enough to discriminate between activities, and could translate well to a visual arts intervention.

The INTERACT measure also is rated from video. The authors noted that they would change the scoring system of the INTERACT for future studies as a sum score is not given, and they felt that item-by-item analysis could increase the risk of false positive results (van Weert et al., 2005). They propose a scale consisting of 'multi-item subscales measuring the same domains' (van Weert et al., 2005, p. 32). This same critique is also mentioned in another study using the INTERACT scale to investigate the physiological and behavioural effects of Snoezelen (van Diepen et al., 2002). Despite this, the INTERACT scale seems suitable to be used to evaluate an intervention. Using video recordings means that a more in depth observation can be made. The only reservation one might have is that by analysing the video content after the sessions, it is doubling the time needed. At 22 items, the scale seems too detailed to be able to be used in live observation, if more than one resident is being observed at one time.

As already suggested, the categories in the OERS may not be sensitive enough or are too broad as four of the six affect states were not often present. Nevertheless, it is a well-used tool. Another limitation of the scale compared to others is the fact that observations are only made of affect, which may be too limiting, a feeling echoed by Fossey et al. (2002) when compared to DCM.

The CEEA was developed with a creative activity in mind and seems simpler to use than some of the other measures. It seems suitable to evaluate the QoL of people with dementia during a psychosocial intervention. However, the measure only gives a broad over-view of the session, rather than time-sampling, for example, so some of the detail might be lost.

The Greater Cincinnati Chapter Well-Being Observation Tool also has the advantage that it was specifically developed for a visual arts programme, but could be used for any psychosocial intervention as shown by it being sensitive to changes within and between different activities.

A recent study critiques the psychometric properties of the Greater Cincinnati Chapter Well-Being Observation Tool (Gross et al., 2013), and perhaps rightly so as it is the first peer-reviewed study to use the tool since development. However, it should be noted that the study itself seems to show some methodological flaws itself. Only one overall rating is made by interns within the MIM session, rather than several ratings made in 10-minute observation windows as intended in the development of the tool. It is unclear whether those making the ratings were the same interns who were facilitating the sessions, meaning that they could be preoccupied by running the session rather than observing the participants. Ratings were then made by staff at the long-term residential facility outside of the sessions 'on or about about each of the same days as the interns . . . at times that were convenient to them' (Gross, et al., 2013, p. 8) implying an inconsistency of time of day that the ratings are made.

A change in the scoring system of the Greater Cincinnati Chapter Well-Being Observation Tool could combat the limitation mentioned in the discussion of training requirements. As the operational definitions are clear, but the extent to which people display each item is more difficult to rate, a change to time sampling whereby each participant is measured for a minute every 8 minutes could simplify it and make the measure more user-friendly.

This tool has the advantage in that it was specifically developed for a visual art intervention. It also seems to give a detailed picture of how the person with dementia experiences the art sessions.

Each observational tool has its own strengths as has been demonstrated in the above review. Most have been shown to have good inter-rater reliability, be rooted in theory and are able to show differences in behaviours within and outside of interventions. None are more detailed than DCM, which seems to have set the standard. However, it is also the most labour intensive and most expensive for training. The PRS and AwareCare tools are sensitive when used for people with severe dementia, but might not be so suitable for a group of mixed ability. The QUIS seems more suited for evaluation of care practices, rather than in evaluating an intervention.

The OERS is a well-used measure but the categories seem a little limited for the rich data that are expected in an evaluation of art sessions. Video recording seems a good option as it means that the data can be analysed without the restriction of predetermined categories, although the resulting Behaviour Observation was noted to be very similar to the already existing MPES. However, using video also doubles the time needed as coding is done at a later date rather than being done in the session. It also opens up the possibility to data being lost without realising until after the session.

The CEAA and Greater Cincinnati Chapter Well-being Observation Tool were developed with a visual art session in mind and the Greater Cincinnati Chapter Well-Being Observation Tool has been shown to be sensitive enough to show differences within sessions and between activities. However, whereas the CEAA gives more of an overview of the session, the Greater Cincinnati Chapter Well-Being Observation Tool gives more detail. It is likely to be a good choice for an arts based evaluation.

It would also seem that a combination of observation in real time coupled with a video recording would be the best solution. This way, the video recordings can be a backup if any data is missed within a session, and inter-rater reliability can be assessed without the need of a second rater within the session, and without the entire onus being on analysis after the event.

Conclusion

No observational measure will capture all of the rich data demonstrated within a visual art intervention. Brooker (2008) suggests that qualitative interviews and observational measures used together complement each other. Therefore, as well as using observational tools to provide quantitative data, it is also advisable to include qualitative interviews with participants, care staff and practitioners/those delivering the intervention to gain a detailed understanding through complimentary methodological perspectives.

This review indicates that based on current research to date, the Greater Cincinnati Chapter Well-Being Observation Tool may be the most appropriate measure to evaluate a psychosocial intervention, and in particular, a visual art programme. Videoing the sessions would add to best practice to help achieve the best assessment of well-being of people with dementia in an art session.

Conflict of interest

The authors declare that there is no conflict of interest.

Funding

This work was supported by National Institute for Social Care and Health Research (NISCHR), Welsh Government, as part of a funded PhD study. They had no role in study design, data collection and analysis, manuscript preparation and revision, or decisions for publication of the current report.

References

- Altman, D. G. (1991). Some common problems in medical research. In: *Practical statistics for medical research* (pp. 396–439). London, England: Chapman & Hall.
- Alzheimer's Disease International (2009). *World Alzheimer's report*. London, England: Alzheimer's Disease International. Retrieved from <http://www.alz.co.uk/research/worldreport/>.
- Alzheimer's Society (2010). My name is not dementia: People with dementia discuss quality of life indicators. London, England: Alzheimer's Society. Retrieved from http://www.alzheimers.org.uk/site/scripts/download_info.php?downloadID=418
- Anderson, K., Bird, MacPherson, S., McDonough, V., & Davis, T. (2011). Findings from a pilot investigation of the effectiveness of a Snoezelen room in residential care: Should we be engaging with our residents more? *Geriatric Nursing*, 32, 166–177.
- Aspland, H., & Gardner, F. (2003). Observational measures of parent-child interaction: An introductory review. *Child and Adolescent Mental Health*, 8, 136–143.
- Baker, R., & Dowling, Z. (1995). *INTERACT*. Dorset: Poole Hospital Research and Development Support Unit.
- Banerjee, S., Smith, S., Lamping, D., Foley, B., Smith, P., & Murray, J. (2004). DEMQOL – evaluation of a new system for measuring quality of life in people with dementia: Validity, reliability and use in regular practice. *Neurobiology of Aging*, 25, S321.
- Baring Foundation. (2011). *Creative homes: How can the arts contribute to quality of life in residential care?* London, England: Baring Foundation. Retrieved from <http://www.baringfoundation.org.uk/CreativeCareHomes.pdf>
- Beard, R. (2011). Art therapies and dementia care: A systematic review. *Dementia*, 11, 633–656.
- Bradford Dementia Group. (2008). *The Bradford well-being profile*. Bradford: University of Bradford.
- Brod, M., Stewart, A. L., Sands, L., & Walton, P. (1999). Conceptualization and measurement of quality of life in dementia: The dementia quality of life instrument (DQoL). *Gerontologist*, 39, 25–35.
- Brooker, D. (1995). Looking at them, looking at me. A review of observational studies into the quality of institutional care for elderly people with dementia. *Journal of Mental Health*, 4, 145–156. DOI:10.1080/09638239550037686.
- Brooker, D. (2005). Dementia Care Mapping: A review of the research literature. *The Gerontologist*, 45, 11–18.
- Brooker, D. (2008). Quality: The perspective of the person with dementia. In M. Downs, & B. Bowers (Eds.), *Excellence in dementia care: Research into practice* (pp. 476–492). Maidenhead: Open University Press.
- Brooker, D., & Duce, L. (2000). Wellbeing and activity in dementia: A comparison of group reminiscence therapy, structured goal-directed group activity and unstructured time. *Aging & Mental Health*, 4, 354–358.
- Brooker, D., & Surr, C. (2006). Dementia Care Mapping (DCM): Initial validation of DCM 8 in UK field trials. *International Journal of Geriatric Psychiatry*, 21, 1018–1025. doi:10.1002/gps.1600.
- Bruce, E. (2000). Looking after well-being: A tool for evaluation. *Journal of Dementia Care*, Nov/Dec, 25–27.
- Clare, L., Whitaker, R., Quinn, C., Jelley, H., Hoare, Z., Woods, B., . . . Wilson, B. (2012). AwareCare: Development and validation of an observational measure of awareness in people with severe

- dementia. *Neuropsychological Rehabilitation: An International Journal*, 22, 113–133. doi:10.1080/09602011.2011.640467.
- Cooper, C., Mukadam, N., Katona, C., Lyketsos, C. G., Ames, D., Rabins, P., . . . Livingstone, G. (2012). Systematic review of the effectiveness of non-pharmacological interventions to improve quality of life of people with dementia. *International Psychogeriatrics*, 24, 856–870.
- Dean, R., Briggs, K., & Lindsay, J. (1993). The DOMus philosophy: A prospective evaluation of two residential units for the elderly mentally ill. *International Journal of Geriatric Psychiatry*, 8, 807–817.
- Dean, R., Proudfoot, R., & Lindsay, J. (1993). The quality of interactions schedule (QUIS): development, reliability and use in the evaluation of two domus units. *International Journal of Geriatric Psychiatry*, 8, 819–826.
- Ettema, T. P., Dröes, R.-M., de Lange, J., Mellenbergh, G. J., & Ribbe, M. W. (2005). A review of quality of life instruments used in dementia. *Quality of Life Research*, 14, 675–686.
- Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-attribute health status classification systems: Health Utilities Index. *PharmoEconomics*, 7, 490–502.
- Feeny, D. H., Furlong, W. J., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., . . . Boyle, M. (2002). Multi-attribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care*, 40, 113–128.
- Fossey, J., Lee, L., & Ballard, C. (2002). Dementia Care Mapping as a research tool for measuring quality of life in care settings: psychometric properties. *International Journal of Geriatric Psychiatry*, 17, 1064–1070.
- Gottlieb-Tanaka, D., Lee, H., & Graf, P. (2008a). *Creative-expressive abilities assessment: A user guide*. Vancouver, British Columbia: ArtScience Press.
- Gottlieb-Tanaka, D., Lee, H. & Graf, P. (2008b). *The CEAA: An instrument for assessing the creative-expressive abilities of individuals with dementia*. ArtScience Press.
- Gross, S. M., Danilova, D., Vandehey, M. A., & Diekoff, G. M. (2013). Creativity and dementia: Does artistic activity affect well-being beyond the art class? *Dementia*. Advance online publication. doi:10.1177/1471301213488899.
- Hadley, C., Brown, S., & Smith, A. (1999). Evaluating interventions for people with severe dementia: Using the positive response schedule. *Aging & Mental Health*, 3, 234–240.
- Harris, P. B. (Ed.) (2002). *The person with Alzheimer's disease: Pathways to understanding the experience*. London and Baltimore: John Hopkins University Press.
- Jarrott, S. E., Gozali, T., & Gigliotti, C. M. (2008). Montessori programming for persons with dementia in the group setting: An analysis of engagement and affect. *Dementia*, 7, 109–125. doi:10.1177/1471301207085370.
- Judge, K. S., Camp, C. J., & Orsulic-Jeras, S. (2000). Use of Montessori-based activities for clients with dementia in adult day care: Effects on engagement. *American Journal of Alzheimer's Disease and Other Dementias*, 15, 42–46.
- Keady, J. (1996). The experience of dementia: A review of the literature and implications for clinical practice. *Journal of Clinical Nursing*, 5, 275–288.
- Kinney, J., & Rentz, C. (2005). Observed well-being among individuals with dementia: Memories in the Making, an art program, versus other structured activity. *American Journal of Alzheimer's Disease and Other Dementias*, 20, 220–227.
- Kitwood, T., & Bredin, K. (1992). A new approach to the evaluation of dementia care. *Journal of Advances in Health and Nursing Care*, 1, 41–60.
- Kline, P. (1979). The meaning of psychometrics. *Psychometrics and Psychology* (pp. 1–25). London: Academic Press Inc.
- Lawton, M. P. (1983). The dimensions of wellbeing. *Experimental Aging Research*, 9, 65–72.
- Lawton, M. P. (1991). A multidimensional view of quality of life in frail elders. In J. E. Lubben, J. C. Rowe, & D. E. Deutchman (Eds.), *The concept and measurement of quality of life* (p. 6). New York: Academic Press.

- Lawton, M. P., Van Haitsman, K., & Klapper, J. (1996). Observed affect in nursing home residents with Alzheimer's disease. *Journal of Gerontology: Psychological Sciences*, *51B*, 3–14.
- Lawton, M. P., Van Haitsman, K., Perkinson, M., & Ruckdeschel, K. (1999). Observed affect and quality of life in dementia: Further affirmations and problems. *Journal of Mental Health and Aging*, *5*, 69–81.
- Logsdon, R. G., Gibbons, L. E., McCurry, S. M., & Teri, L. (1999). Quality of life in Alzheimer's disease: Patient and caregiver reports. *Journal of Mental Health and Aging*, *5*, 21–32.
- MacPherson, S., Bird, M., Anderson, K., Davis, T., & Blair, A. (2009). An art gallery access programme for people with dementia: 'You do it for the moment'. *Aging & Mental Health*, *13*, 744–752.
- Magaziner, J., Simonsick, E. M., Kasher, T. M., & Hebel, J. R. (1988). Patient-proxy response comparability on measures of patient health and functional status. *Journal of Clinical Epidemiology*, *41*, 1065–1074.
- Moniz-Cook, E., Vernooij-Dassen, M., Woods, R., Verhey, F., Chattat, R., Vugt, D. E., Orrell, M. for The INTERDEM Group. (2008). A European consensus on outcome measures for psychosocial intervention research in dementia care. *Aging & Mental Health*, *12*, 14–29.
- Olazarán, J., Reisberg, B., Clare, L., Cruz, I., Peña-Casanova, J., del Ser, T., . . . Muniz, R. (2010). Nonpharmacological therapies in Alzheimer's disease: A systematic review of efficacy. *Dementia and Geriatric Cognitive Disorders*, *30*, 161–178.
- Perrin, T. (1997). The positive response schedule for severe dementia. *Aging & Mental Health*, *1*, 184–191.
- Phillips, L. J., Reid-Arndt, S. A., & Pak, Y. (2010). Effects of a creative expression intervention on emotions, communication, and quality of life in persons with dementia. *Nursing Research*, *59*, 417–425.
- Robins, P. V., Kasper, J. D., Kleinman, L., Black, B. S., & Patrick, D. L. (1999). Concepts and methods in the development of the ADRQL: An instrument for assessing health-related quality of life in persons with Alzheimer's disease. *Journal of Mental Health and Aging*, *5*, 33–48.
- Ready, R. E., Ott, B. R., Grace, J., & Fernandez, I. (2002). The Cornell-Brown scale for quality of life in dementia. *Alzheimer's Disease and Associated Disorders*, *16*, 109–115.
- Rentz, C. (2002). Memories in the Making[®]: Outcome-based evaluation of an art program for individuals with dementing illnesses. *American Journal of Alzheimer's Disease and Other Dementias*, *17*, 175–181.
- Sabat, S. R. (2001). *Life through a tangled veil: The experience of Alzheimer's disease*. Oxford: Blackwell.
- Salisbury, K., Algar, K., & Windle, G. (2011). Arts programmes and quality of life for people with dementia – a review. *Journal of Dementia Care*, *19*, 33–37.
- Sheard, D. (2008). *Enabling: Quality of life – an evaluation tool*. London, England: Alzheimer's Society.
- Sloane, P. D., Brooker, D., Cohen, L., Douglass, C., Edelman, P., Fulton, B. R., . . . Zimmerman, S. (2007). Dementia care mapping as a research tool. *International Journal of Geriatric Psychiatry*, *22*, 580–589.
- Sterin, G. J. (2002). Essay on a word. A lived experience of Alzheimer's disease. *Dementia*, *1*, 7–10.
- Streiner, D. L., & Norman, G. R. (2003). *Health Measurement Scales: A practical guide to their development and use*. Oxford: Oxford University Press.
- Terada, S., Ishizu, H., Fujisawa, Y., Fujita, D., Yokota, O., Nakashima, H., . . . Kuroda, S. (2002). Development and evaluation of a health-related quality of life questionnaire for the elderly with dementia in Japan. *International Journal of Geriatric Psychiatry*, *17*, 851–858.
- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D., Knol, D. L., Dekker, J., . . . de Vet, H. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*, 34–42.

- van Diepen, E., Baillon, S. F., Redman, J., Rooke, N., Spencer, D. A., & Prettyman, R. (2002). A pilot study of the physiological and behavioural effects of Snoezelen in dementia. *British Journal of Occupational Therapy*, 65, 61–66.
- van Weert, J. C. M., van Dulmen, A. M., Spreeuwenberg, P. M. M., Ribbe, M. W., & Bensing, J. M. (2005). Behavioural and mood effects of snoezelen integrated into 24 hour care. *Journal of the American Geriatrics Society*, 53, 24–33.
- Weiner, M. F., Martin-Cook, K., Svetlik, D. A., Saine, K., Foster, B., & Fontaine, C. S. (2000). The quality of life in late-stage dementia (QUALID) scale. *American Journal of the American Medical Medical Directors Association*, 1, 114–116.
- World Health Organization. (1995). The World Health Organization quality of life assessment (WHOQOL): Position paper from the World Health Organization. *Social Science and Medicine*, 41, 1403–1409.
- World Health Organization. (2004). *Ageing and Health Technical Report Volume 5: A glossary of terms for community health care and services for older persons*. World Health Organisation Centre for Health Development, Geneva. Retrieved from http://whqlibdoc.who.int/wkc/2004/WHO_WKC_Tech.Ser._04.2.pdf

Katherine Algar works at the Dementia Services Development Centre, Wales, and for NEURODEM Cymru at Bangor University. This paper forms part of her PhD study looking at the benefits of visual art programmes for care home residents with dementia.

Robert T Woods is a Professor of Clinical Psychology of Older People at Bangor University, Director of NEURODEM Cymru and co-director of the Dementia Services Development Centre Wales. His research interests include the evaluation of psychosocial interventions in dementia care.

Gill Windle is a Senior Research Fellow at the Dementia Services Development Centre, Wales at Bangor University. Her research interests include mental health and resilience in later life; creativity and ageing; and the interplay between the individual and their physical, social and environmental contexts on well-being and quality of life.