Momentary well-being in dementia

Observational assessment of psychosocial interventions and music therapy

Kristine Gustavsen Madsø

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2022

UNIVERSITY OF BERGEN

Momentary well-being in dementia

Observational assessment of psychosocial interventions and music therapy

Kristine Gustavsen Madsø

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 04.03.2022

Year: 2022

Title: Momentary well-being in dementia

Name: Kristine Gustavsen Madsø

Print: Skipnes Kommunikasjon / University of Bergen

Scientific environment

This thesis was part of a collaborative project between the Department of Clinical Psychology at the Faculty of psychology (UiB) and NKS Olaviken Gerontopsychiatric Hospital. The PhD training was completed at the Graduate School of Clinical and Developmental Psychology (CDP) at the Faculty of Psychology (UiB) in parallel with the candidate's education as a specialist in Clinical Gerontopsychology at NKS Olaviken Gerontopsychiatric Hospital. The research group affiliation was Bergen Research Group for Innovation, Growth, Health and Technology (BRIGHT).

The research was funded through a PhD grant from the Faculty of Psychology (UiB), with supplementary funding from Meltzer Research Foundation (under grant number 811305) and POLYFON Knowledge Cluster of Music Therapy (under grant number 811425) via NKS Olaviken Gerontopsychiatric Hospitals membership in the cluster. NKS Olaviken Gerontopsychiatric Hospital also supported recruitment of participants to the studies.

Alrek
helseklynge

NKS OLAVIKEN
ALDERSPSYKIATRISKE SYKEHUS

L. MELTZERS HØYSKOLEFOND

polyfon | kunnskapsklynge
for musikkterapi

Acknowledgements

First, I wish to express my gratitude to the participants of the research projects who openly shared their invaluable knowledge and experiences with us. In particular, the courage to allow a camera to document such intimate moments, is admirable. Several of you have expressed a motivation of sharing your experiences of living with dementia to help other people in the future. It is my sincere hope that this thesis will be a useful contribution.

To my dear main-supervisor Inger Hilde Nordhus, your engagement in gerontology have been very inspiring. You are such a warm and knowledgeable person. I am so glad to have received your kind and supportive guidance over these years. Co-supervisor Minna Hynninen, you are the kind of person that make things happen. The way you have supported the research project at NKS Olaviken Gerontopsychiatric Hospital have been vital. I am looking forward to seeing, and hopefully contributing to, the research or development projects you will make happen in the future. Co-supervisor Helge Molde, thank you for sharing your competence and passion for statistical methods and psychology. Somehow you always manage to instantly e-mail an article perfectly explaining any small or big issue encountered. You even made me use R. Also, the collaboration with co-authors Nancy Pachana and Elisabeth Flo-Groeneboom have been greatly appreciated.

I am so thankful to the wonderful colleagues and fellow PhD-candidates at the Department of Clinical Psychology and at NKS Olaviken Gerontopsychiatric Hospital. I wish to thank all the incredible music therapist I have collaborated with and learned so much from during the last years: Solgunn Knardal, Sunniva Ulstein Kayser, Gerd Karine Lunde Thorsvik, Johan Klætte, and Camilla Hay. The priceless guidance from librarian Kjersti Aksnes-Hopland at the University Library is something I am very thankful for. I also owe gratitude to students Sofie Holt Korneliussen and Marte Muri for their thorough work with the video-coding and the music therapy project.

I feel very grateful for receiving the funding and support from Meltzer Research Fund, POLYFON, NKS Olaviken Gerontopsychiatric Hospital, and the University of Bergen. Specifically, Frode Wikne, director of NKS Olaviken Gerontopsychiatric Hospital, your support of the research project has been much appreciated.

My dear friends, family, and my parents; thank you so much for always supporting me and believing in me when I doubted. Selma Næss, my dear grandmother, thank you for contributing with your thoughts, presence, and dinners. And finally, Bård, during these years you have been solid as a rock. Thank you for your patience. I was "engaged" with this thesis before we married, but even when my mind was absent you always had my heart. Finally, I am all yours.

Sincerely,
Kristine

"If I am no longer a woman, why do I still feel one? If no longer worth holding, why do I crave it? If no longer sensual, why do I enjoy the soft texture of silk against my skin? If no longer sensitive, why do moving lyric songs strike a responsive chord in me? My every molecule seems to scream out that I do, indeed, exist, and that my existence must be valued by someone! Without someone to walk this labyrinth by my side, without the touch of a fellow traveler who understands my need of self-worth, how can I endure the rest of this uncharted journey?" (McGowin, 1993, pp. 123-124, *Living in the Labyrinth*)

"Once one has seen such responses one knows that there is still a self to be called upon, even if music and only music, can to the calling…to those who are lost in dementia…music is no luxury to them, but a necessity, and can have a power beyond anything else to restore them to themselves, and to others, at least for a while." (Oliver Sacks, 2007, pp. 346-347, *Musicophilia: Tales of Music and the Brain*)

Abstract

Background: Dementia is an increasingly prevalent condition world-wide. A cure is yet to be found. The adverse behavioral and psychological symptoms accompanying dementia are best approached through non-pharmacological multi-component interventions. However, more than management of adverse symptoms is required to live well with dementia. Recent initiatives call for assessing interventions within the frame of positive psychology. People with dementia want interventions promoting well-being, facilitating continued relationship closeness, and supporting independency. Such outcomes guided the three research projects in this thesis.

Music therapy has so far demonstrated its potential as a beneficial approach for people living with dementia. Studies have primarily been conducted in long-term care and in populations with a severe dementia. However, most people with dementia live at home. A central aim of this thesis was to investigate music therapy for the home-dwelling population within a positive psychological framework, in particular assessing well-being and sociable interaction. Most dementias have a progressive course, which complicates a valid assessment of interventions over long time intervals. Assessment over short intervals in real-time appears advantageous to detect clinically important changes in well-being. This is possible through ecologic momentary assessment and observational methods. This thesis aimed to provide a systematic overview of available observational momentary well-being measures relevant for dementia research. No observational instruments were available for comparing well-being during music therapy and other interventions validly. Thus, the third aim of this thesis was to develop an instrument sensitive for detecting clinically significant changes in well-being over short time-lags, comparing music-therapy to other interventions.

Methods: Paper 1 is a systematic review of observational measures assessing well-being over small time-lags. The review is based on the Consensus based Standards for selection of health Measurement Instruments (COSMIN), assessing central measurement properties, feasibility, and interpretability. Paper 2 is a single-case study with a bi-phasic AB design replicated three times per participant. It included 11 participants and over 2000 observations. The study compared music therapy and regular social interaction for home-dwelling people with dementia and a close caregiver. The intervention was assessed with (1) observational measures investigating sociable interaction and well-being investigated in meta-analyses; (2) self-reported emotions for each session; (3) long-term well-being, neuropsychiatric symptoms, and caregiver burden before and after the 10-week intervention. Paper 3 describes the development of the Observed Well-being in Living with dementia Scale, an instrument assessing well-being during music therapy and social interaction. The instrument was developed through an iterative mixed methods approach including fieldwork, field-testing, psychometric properties assessment, focus group assessment from experts, and revisions and final field-testing.

Results: Paper 1 identified 22 instruments assessing well-being through observation. Content validity was the best documented measurement property. No instruments assessed responsiveness and very few assessed measurement invariance. Paper 2 found music therapy to promote individual well-being and sociable interactions towards significant others. Self-rated positive emotions were increased after the sessions. From pre- to post-sessions, neuropsychiatric symptoms were lowered. Long-term well-being and caregiver burden were stable. Paper 3 operationalized well-being as consisting of the eight items “attention”, “initiative/response”, “calm/relaxed”, “happiness”, “enjoyment”, “express identity”, “mastery”, and “relationship”. Content validity was adequate, reliability measures were good, and preliminary psychometric properties were promising.

Conclusion: This thesis adds to the current knowledge of well-being for people with dementia. It provides a structured evaluation of the current observational well-being instruments, their measurement properties, and their clinical utility. The intervention study found music therapy to have a potential to increase well-being and positive emotions in the moment and calls for replication. Additionally, the increase in sociable interaction can potentially facilitate relationship closeness. The observational instrument developed needs further assessment in larger samples and other contexts.

List of Publications

Paper 1: Madsø, K. G., Flo-Groeneboom, E., Pachana, N. A., & Nordhus, I. H. (*Manuscript under review*). Assessing momentary well-being in people living with dementia: A systematic review of observational instruments.

Paper 2: Madsø, K. G., Molde, H, Hynninen, K. M., & Nordhus, I. H. (2021). Observing music therapy in dementia: Repeated single-case studies assessing well-being and sociable interaction. *Clinical Gerontologist*. Advance online publication. https://doi.org/10.1080/07317115.2021.1978121

Paper 3: Madsø, K. G., Pachana, N., & Nordhus, I. H (*Submitted manuscript*). Development of the Observable Well-being in Living with Dementia-Scale (OWLS): a measure for well-being during psychosocial and creative interventions.

Abbreviations

ACT – Activity in Context and Time

AD – Alzheimer's disease

AER – The Apparent Emotion Rating Scale

BEAM – Behavior, Engagement and Affect Measure

BPSD – Behavioral and psychological symptoms of dementia

CDR – Clinical dementia rating

COSMIN - Consensus based Standards for selection of health Measurement INstruments

DCM - 8 – Dementia care mapping – version 8

ELICSE - Ethographic and Laban Inspired Coding System of Engagement

EMFACS/FACS - (Emotion) Facial Action Coding System

EMODEB - Evidence-Based Model of Engagement-Related-Behavior

EPWDS - Engagement of a Person with Dementia Scale

FTD – Frontotemporal dementia

GCWBT - Greater Cincinnati Chapter Well-Being Observational Tool

GRADE - Grading of Recommendations Assessment, Development and Evaluation

LRR – Log Response Ratio

DLB – Dementia with Lewy bodies

MAX – The Maximally Discriminative Facial Movement Coding System

MEDLO-tool – Maastricht Electronic Daily Life Observational tool

MiDAS – Music in Dementia Assessment Scales

MPES – Menorah Park Engagement Scale

MTED – Music Therapy Engagement Scale

NICE – National Institute of Health and Care Excellence

NPI-Q – Neuropsychiatric Inventory-Questionnaire

OERS – Observed Emotion Rating Scale

ODAS – Observable Displays of Affect Scale

OME – Observational measure of Engagement

OWLS – Observed Well-being in Living with dementia Scale

QUALIDEM-ILA – QUALIDEM for Intensive Longitudinal Assessment
QoL – Quality of life
QoL-AD – Quality of Live in Alzheimer's Disease
PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO – International prospective register of systematic reviews
PRS – Positive Response Schedule
REK – Regional Committees for Medical and Health Research Ethics in Norway
RoB – Risk of Bias
RSS – Relative Stress Scale
SCRIBE – Single-Case Reporting Guideline in Behavioral Interventions
SM-GCWBT – Scripps Modified Greater Cincinnati Chapter Well-Being Observational Tool
VAMS – Visual Analogue Mood Scale
VC-IOE – Video Coding – Incorporating Observed Emotion
VD – Vascular dementia
VNVIS-CR – Verbal and Nonverbal Sociable Interaction Scale- Care Receiver
WHO – World Health Association

Contents

1. Introduction

1.1 Purpose of the thesis

The prevalence of dementia is increasing world-wide, and the research society has so far not succeeded in finding a cure (Prince et al., 2016). Meanwhile, it is also important to understand how dementia may be lived with well and which interventions best can facilitate this throughout dementia progression. Music therapy is one of several interventions that have gained focus over the last two decades, with the potential to alleviate symptoms and mood in dementia. Still, it is necessary to add to the current knowledge base about the potential positive responses during music therapy for people with dementia (Dowson et al., 2019; van der Steen et al., 2018). Thus, the initial objective of this thesis was to investigate changes during a music therapy intervention through a positive psychology framework (Stoner et al., 2019). The assessment of momentary changes of well-being, and sociable behavior when interacting with significant others, were our main outcomes. While searching for an adequate observational well-being measure, a need to systematize the available observational measures became apparent. Consequently, an additional objective for this thesis was to conduct a systematic review on observational well-being measures and complement other reviews and initiatives recommending the use of positive outcomes in dementia research (Clarke et al., 2020; Dröes et al., 2016; Harrison et al., 2016; Reilly et al., 2020; Stoner et al., 2019; Øksnebjerg et al., 2018).

While music therapy may bear promise for promoting well-being in dementia, specific challenges with the validity and sensitivity of observational outcome measures for assessing music interventions were evident. No feasible observational instruments for comparing well-being during music therapy to other psychosocial interventions were found. The observational video-data gathered in the music therapy study was the starting point of developing such an instrument.

1.2 Structure of the thesis

The introduction will briefly review what dementia is and the related societal, individual, and relational consequences. Next, the focus on living well with dementia is described using a biopsychosocial perspective. To define clinically relevant outcomes for psychosocial interventions key needs as described by people living with dementia are presented. As well-being is central here, this construct is defined, and factors associated with well-being are outlined. Further, interventional approaches targeting well-being in dementia are summarized. Music therapy is elaborated as an intervention capable of fulfilling some of these the central needs. Several reviews have identified methodological issues complicating the detection of treatment effects in psychosocial interventions and music therapy. These issues are described, followed by suggestions of using momentary assessment and observation to solve some of these pertinent issues.

After the introduction, I describe the specific aims and research questions for each paper. Because this thesis utilizes both qualitative and quantitative methods, a section elaborating on the use of mixed methods is provided. This section is followed by an overview and justification of the specific methods chosen to investigate the questions and aims. Then, the results are presented, and discussed in relation to existing research, methodological strengths and limitations, ethical considerations, and the implications of the results for further research and clinical practice.

1.3 Dementia: diagnostic features, prevalence, and consequences

Dementia refers to a clinical syndrome associated with degeneration of or damage to brain tissue of multiple etiological causes. Current pharmacologic interventions can, at best, slow down the progression of dementia, but no curative treatment has yet been discovered (Winblad et al., 2016). It is estimated that between 4.6 – 8.7% of the population over 60 years old will develop dementia (Prince et al., 2015). Alzheimer's dementia (AD) is the most common subtype of dementia, followed by Vascular dementia (VD), Dementia with Lewy bodies (DLB) and dementia with mixed etiologies. The least common is Frontotemporal dementia (FTD) and other dementias related to brain injury, infections, or alcohol use (Livingston et al., 2017). In this dissertation, the concept of "dementia" refers to the three most common forms unless otherwise specified.

It is estimated that around 50 million people live with dementia globally, anticipating a rise to 131 million people by 2050 (Livingston et al., 2017; Prince et al., 2016). Although some evidence points to a lower incidence in developed countries in recent years (Wu et al., 2017), the global prevalence is still rising in conjunction with the aging population of the world (Alzheimer's Disease International & Patterson, 2018). In Norway, the current number of people living with dementia is estimated to be over 100 000 in a population of 5.4 million people (Gjøra et al., 2021).

Table 1 presents the core diagnostic features of dementia in the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-V; American Psychiatric Association (APA), 2013) and the International Statistical Classification of Diseases and Health-Problems (5th. ed; ICD-10; World Health Organisation (WHO), 2016).

While the diagnostic features vary slightly between ICD-10 and DCM-V, both descriptions imply living with dementia has various consequences on individual, interpersonal, and societal levels.

Table 1 – Core diagnostic features of dementia and major neurocognitive disorder

ICD-10 definition of dementia (WHO, 2016, p 43 – 44)	**DSM-V criteria for major neurocognitive disorder (APA, 2013)**
Dementia (F00-F03) is a syndrome due to disease of the brain, usually of a chronic or progressive nature, in which there is disturbance of multiple higher cortical functions, including memory, thinking, orientation, comprehension, calculation, learning capacity, language, and judgement. Consciousness is not clouded. The impairments of cognitive function are commonly accompanied, and occasionally preceded, by deterioration in emotional control, social behavior, or motivation.	A. Evidence of significant cognitive decline from a previous level of performance in one or more cognitive domains (complex attention, executive function, learning and memory, language, perceptual–motor, or social cognition) based on: 1. Concern of the individual, a knowledgeable informant, or the clinician that there has been a significant decline in cognitive function; and 2. A substantial impairment in cognitive performance, preferably documented by standardized neuropsychological testing or, in its absence, another quantified clinical assessment.
	B. The cognitive deficits interfere with independence in everyday activities (that is, at a minimum, requiring assistance with complex instrumental activities of daily living such as paying bills or managing medications)
	C. The cognitive deficits do not occur exclusively in the context of a delirium.
	D. The cognitive deficits are not better explained by another mental disorder. Specify: - Without behavioral disturbance: if the cognitive disturbance is not accompanied by any clinically significant behavioral disturbance - With behavioral disturbance (specify disturbance): if the cognitive disturbance is accompanied by a clinically significant behavioral disturbance (for example, psychotic symptoms, mood disturbance, agitation, apathy, or other behavioral symptoms). For example, major depressive disorder or schizophrenia

1.3.1 The societal consequences of dementia

At a societal level, the increase of dependency following dementia has economic consequences. Throughout the progression of dementia, comprehensive caregiving is necessary, contributing to high demands on informal caregivers or leading to institutionalization. Globally, 85% of the US$818 billion estimated as the total economic expenses related to dementia is associated with family and social care (Livingston et al., 2017; Prince et al., 2016).

1.3.2 The individual consequences of dementia

Although dementia progress will vary between individuals, a development from mild and moderate to severe stages during a 5 – 15-year period is expected (Livingston et al., 2017). The progression of neurodegeneration or damage is always interfering with the ability to perform daily activities. A common approach to staging the dementia severity is through the Clinical Dementia Rating assessment (CDR), where functioning is assessed for the six assets ‘memory,’ ‘orientation,’ ‘judgment and problem solving,’ ‘community affairs,’ ‘home and hobbies,’ and ‘personal care’ (Hughes et al., 1982; Morris, 1993). As the dementia progresses, impairment of the abovementioned functions raises the risk on a broad range of issues such as the threat of nutritional deficits, lowered adherence to medication, issues of commuting safety, and social isolation (Livingston et al., 2017).

The individual emotional distress of living with dementia was recently elaborated in a comprehensive systematic review including 121 quantitative and qualitative publications reporting from the perspective of people living with dementia (Petty et al., 2018). The most common distressing experiences were being fearful and feeling lonely. Emotional distress was clustered into five main themes: describing insecurity about oneself, awareness of being in a stigmatizing social position, lack of relational needs being met, worries about the future, struggling to manage demanding physical environments, and experiencing perceptional disturbances (Petty et al., 2018). In the *early* stages of dementia, commonly expressed psychological distress include loss of control and autonomy, impaired feelings of self-worth, and loss of former role-functions (Bjørkløf et al., 2019; Von Kutzleben et al., 2012). Withdrawal from social activities is a common reaction to these distressing experiences (Miranda-Castillo et al., 2013). Social isolation is eventually a risk factor for more rapid cognitive decline and the development of depression (Livingston et al., 2017).

Additionally, neuropsychiatric symptoms or “behavioral and psychological symptoms of dementia” (BPSD) often accompany dementia progression (Kales et al., 2015; Kolanowski et al., 2017; Lyketsos et al., 2011; Nowrangi et al., 2015). In this thesis, these terms are used synonymously. According to the most used assessment scale of neuropsychiatric symptoms (The Neuropsychiatric Inventory-Questionnaire), these symptoms are anxiety, apathy, depression, sleep disturbances, appetite disturbances, euphoria, delusions, disinhibition, hallucinations, irritability, aggression/agitation, and aberrant motor behavior (Kaufer et al., 2000).

Longitudinal studies show that over five years, 97% of people living with dementia experienced at least one BPSD (Steinberg et al., 2008), with apathy (43 – 59%), depression (37 – 41%), and anxiety (37 – 41%) as the most common experiences (Leung et al., 2021; Steinberg et al., 2008).

1.3.3 The relational consequences of dementia

The increased dependency following the progression of dementia, implies that the frequent individual consequences presented above have interpersonal consequences (WHO, 2017). The changes will ultimately challenge the relationship between the person living with dementia and their significant others and can lead to a decline in intimacy and reciprocity, and complicate communication (Ablitt et al., 2009).

Before receiving a diagnosis of dementia, the cognitive, social, and behavioral changes are likely to be misunderstood by others. If significant others evaluate these changes as primarily negative personal intentions or characteristics, this evaluation may lead to feelings of frustration or humiliation and result in impaired self-confidence in the individual with dementia (Ablitt et al., 2009; Bjørkløf et al., 2019; Harding et al., 2019). After a dementia diagnosis, many people living with dementia and their informal caregivers describe a wish to preserve a close relationship (Graham & Bassett, 2006). Although informal caregivers also report that caregiving may be emotionally rewarding, it is very common to experience it as overwhelming and exhausting (Wadham et al., 2016). Some caregivers describe increasing feelings of closeness and mutual affection after dementia onset (Ablitt et al., 2009). However, the presence of BPSD is associated with weakened relationship mutuality and increased caregiver burden (Cheng, 2017). Apathy is associated with a lower degree of well-being, more depression, and a self-reported caregiver burden (Feast et al., 2016; Kolanowski et al., 2017). Damages and/or neurodegeneration in the frontal areas of the brain accompany some types of dementia, cause particularly burdensome symptoms such as lack of inhibition, and inappropriate or aggressive behavior. Visual and auditory hallucinations accompany other disease courses. Common topics of delusions include persecution, theft, or infidelity. All these more severe disruptive behaviors significantly affect social interactions and contribute to a lack of attachment between people with dementia and their family caregivers (Cheng, 2017).

In high-income countries, about two-thirds of people with dementia live at home, supported by close family and/or professional caregivers. In low to middle-income countries, over 90% live at home (Prince et al., 2015). The Norwegian Dementia Strategy 2025, published by The Norwegian Ministry of Health, has a strong focus on providing support to enable people with dementia to live at home for as long as possible (Helsedirektoratet, 2019). However, the hours of informal care provided by family or friends of people with dementia living at home can range from 30 – 100, increasing as the dementia progress (Ydstebø et al., 2020). This statistic reveals that living with dementia also can have a far-reaching impact on family caregivers' lives.

In sum, dementia has severe personal, relational, and societal consequences. Adding the high prevalence to the equation, it is clear why WHO defines dementia as a public health priority (WHO, 2017). Recognizing the emotional pain, frustration and suffering that follows dementia, we need to focus on mechanisms and interventions to alleviate this burden in the window of what is achievable (Bartlett et al., 2017; Spector & Orrell, 2010).

1.4 The focus of living well with dementia

Research on psychosocial interventions in dementia has been criticized for being based on methods best suited for drug-development research, mainly focusing on *impairment* while omitting aspects that may be more meaningful for the target group (Gaugler et al., 2019; O'Rourke, Fraser, et al., 2015). Several initiatives now call for changing the focus of dementia intervention research from management of BPSD or cognitive deficits only, towards including a positive psychology framework where factors contributing to resilience, well-being, QoL, and positive relationships are identified (Gaugler et al., 2019).

Over the last decades, it has been acknowledged in research and care that it is possible to live well with dementia despite its negative consequences (Kwasky et al., 2010; Livingston et al., 2017; Martyr et al., 2018). Living well with dementia includes receiving support to remain as active and independent as possible, to access and use one's capacities and resources, receive individualized care that can intervene when BPSD are troubling, optimized pharmacological treatment, help with comorbid health issues, technology to meet cognitive needs, and sufficient support for family caregivers (Alzheimer's Society, 2021; Kolanowski et al., 2018; Kwasky et al., 2010; Livingston et al., 2017; Woods, 2012). Furthermore, people with dementia are not only passive receivers of support, but they also actively strive to adapt to and compensate for the cognitive and functional impairment (Bjørkløf et al., 2019; Von Kutzleben et al., 2012).

Thus, supporting people to live well with dementia calls for understanding the condition from a broader perspective than the biomedical model (Lord et al., 2020). For this purpose, this thesis utilizes a biopsychosocial conception (Spector & Orrell, 2010).

1.4.1 Dementia in a biopsychosocial perspective

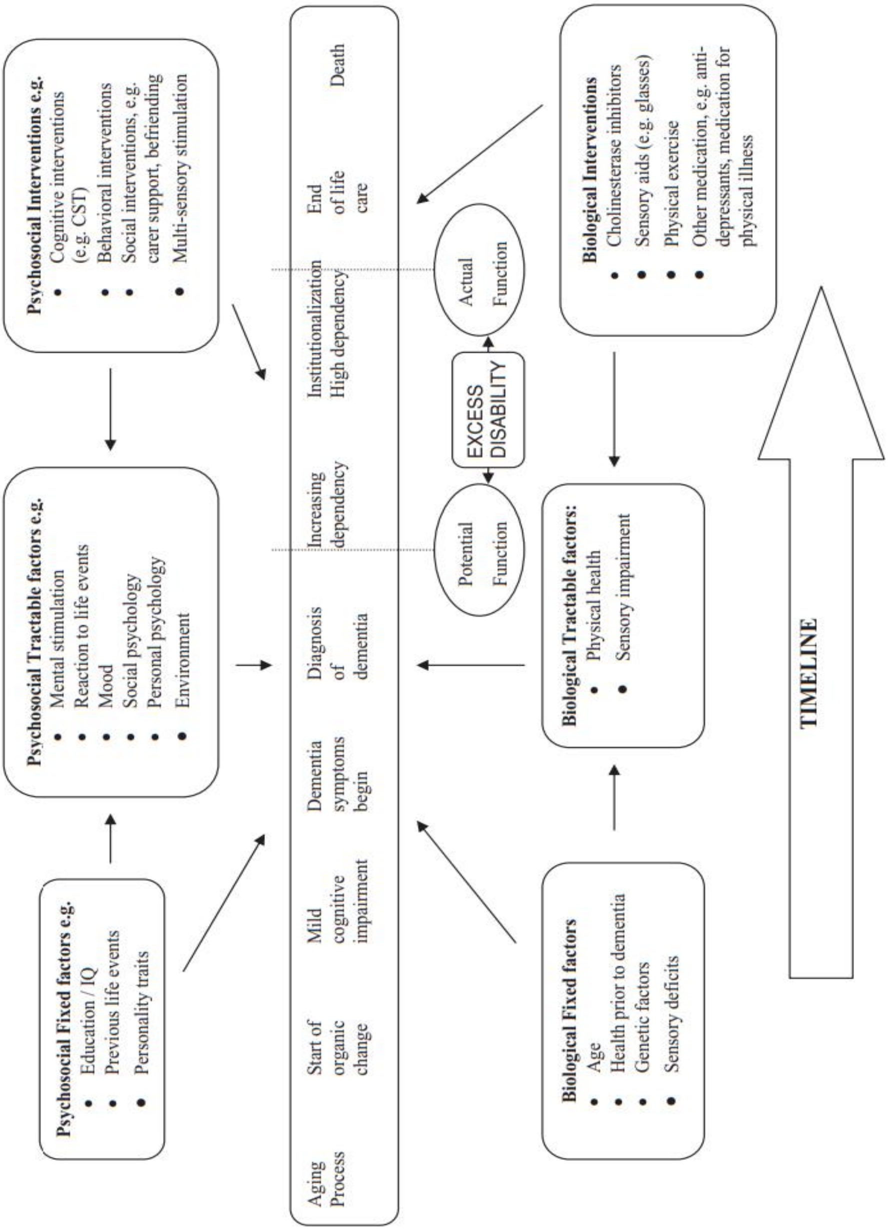
The person-centered dementia model by Kitwood (1997), represents one of the most influential approaches placing dementia in a biopsychosocial as opposed to a biomedical perspective (Lord et al., 2020). Spector and Orrell (2010) elaborated this view in a model as shown in Figure 1 below. Their model can guide development and understanding of dementia interventions. According to this model, some fixed and some amenable factors influence the expression and development of dementia-related symptoms. Thus, several of the manifestations of dementia are *modifiable* through interventions tailored to a person's specific physical, psychological, social, and individual needs (Livingston et al., 2017). Kitwood (1997) addressed the relationship between psychosocial and neurological factors. Harmful mechanisms in the environment people with dementia live in are described as "*malignant social psychology*", and include interactions that for example infantilize, ignore, or de-humanize people (Kitwood, 1997; p 45).

The model of Spector and Orrell (2010) illustrates how the fixed and tractable factors may have a psychological, social/environmental, and biological origin – and thus, offers a guide for more targeted clinical interventions. The dotted lines in the figure capture the interval between *potential* and *actual* function. The functional level of an individual is influenced by their fixed factors *and* modifiable by the tractable factors to which one is exposed. This process opens up an adaptable room of *ability* or *excess disability* dependent on the factors, interventions, or resources one can access. The objective for psychosocial and biological interventions is to focus on maximizing the function of the individual within these boundaries (Spector & Orrell, 2010).

Excess disability can be triggered in interactions with the environment. Examples of excess disability is higher dependency or more adverse emotional experiences than the underlying neurological impairment can account for (Woods, 1999). BPSD can also represent communication of unmet needs (Kitwood, 1997; Kolanowski, 1999), originating from untreated pain, infections or other physiological conditions, lack of social interactions, negative communication patterns, or lack of engaging activities (Kales et al., 2015; Livingston et al., 2017).

This thesis will add to the existing knowledge base about the means and measures of living well with dementia that has developed over recent decades. Addressing what is considered important from the perspective of the individual is also important to deliver sufficient and individualized care (Kolanowski et al., 2018; Livingston et al., 2017; Spector & Orrell, 2010). Thus, the next section gives an overview of key needs and issues described by people living with dementia.

Figure 1 – The biopsychosocial model of dementia



Source: Spector & Orrell (2010), p 960. Reprinted with permission from the publisher (Rightslink® by Copyright Clearance Center)

1.4.2 Key needs described by people living with dementia

Several key needs have been identified as important in order to live well by people with dementia. Home-dwelling people report a need to 'come to terms' with their condition and maintain normality (Bjørkløf et al., 2019; Von Kutzleben et al., 2012), to experience autonomy, empowerment, respect, and confidence (Bjørkløf et al., 2019; Reilly et al., 2020; Von Kutzleben et al., 2012; Øksnebjerg et al., 2018), to experience close relationships, feeling valued, and being able to communicate (Reilly et al., 2020; Von Kutzleben et al., 2012), and to receive sufficient support in practical domains such as hygiene, comfort, prevent falling, aids for vision and hearing (Reilly et al., 2020; Von Kutzleben et al., 2012; Øksnebjerg et al., 2018). Additionally, the most common *unmet needs* reported by people living with dementia are lack of support for psychological distress, meaningful activities, and social company (Miranda-Castillo et al., 2013). These needs define relevant outcome measures for dementia interventions that aim to positively affect the lives of people living with dementia. Living well with dementia covers more than well-being or QoL per se, though individual well-being is an essential part of it. Thus, it is relevant to give an overview of how well-being is defined and measured in dementia research.

1.4.3 Defining well-being in dementia

Well-being may be operationalized as wealth and health at a national or population level, through individual biomarkers or physiological measures, or reflect the subjective evaluation or appraisal of an individual (Burns, 2017; Ferring & Boll, 2010). This thesis focuses on psychological aspects of well-being.

Lawton (1994, 1997) and Kitwood (1997) pioneered the approach of understanding QoL and well-being from the perspectives of people living with dementia. Their classical texts apply a person-centered focus that is now standard in modern dementia research and theory.

Clarke et al. (2020) suggest defining well-being in dementia as "*a state of equilibrium that, when achieved, gives rise to positive psychological health*" (p 9; Clarke et al., 2020; based on Dodge et al., 2012). They propose a conceptual model of well-being that incorporates the classical definitions of subjective well-being (Diener, 1984), social well-being (Kitwood, 1997), and psychological well-being (Ryff, 1989). The well-being model of Clarke et al. (2020) includes life satisfaction as an overarching category that contains these three well-being domains. While *emotional* well-being includes positive emotions, *social* well-being includes relational aspects such as connection and belonging. *Psychological* well-being includes a positive sense of self, making meaning while experiencing the challenges of dementia, and retaining agency and autonomy (Clarke et al., 2020). The definition of Clarke et al. (2020) relies on qualitative inquiries of people with dementia. It appears to represent the construct from the view of the target group adequately (Brod et al., 2009).

QoL and well-being have frequently been used synonymously in the dementia literature, but QoL is traditionally a broader multi-dimensional construct (Bowling et al., 2015). QoL and well-being have been conceptualized in multiple ways in dementia research (Missotten et al., 2016). The interchangeable use of the constructs makes it complicated to review research on well-being and QoL separately. However, all the conceptualizations mentioned above describe the constructs as relative to the perception of the individual, which implies that higher levels of well-being or QoL should be obtainable despite living with a chronic condition (Young et al., 2009).

Well-being and QoL have frequently been defined operationally as *absence* or *decrease* in depression, anxiety, and BPSD in intervention research for dementia (Bowling et al., 2015). This conceptualization is not necessarily accurate, but until recently, assessing well-being or QoL specifically were omitted in most dementia intervention studies (Harrison et al., 2016). A systematic review of factors associated with QoL in persons with dementia found only moderate associations with symptoms of depression and BPSD (Martyr et al., 2018). It seems that QoL and well-being is conceptually different from the absence of adverse symptoms. Well-being and ill-being may even co-occur and be modified by unrelated factors (Bowling et al., 2015; Kolanowski et al., 2017; Kwasky et al., 2010; Lawton et al., 1996; Martyr et al., 2018).

1.4.4 Factors associated with well-being in dementia

Meta-analyses, meta-syntheses, and systematic narrative reviews assessing factors associated with well-being and QoL highlight the differences between factors exposed in self-reported versus proxy-rated data. For the self-reported data, factors associated with well-being and QoL in dementia are having good relationships with both family and professional caregivers, religious beliefs or spirituality, social engagement, perceptions of good health, feelings of belonging and connection, autonomy, agency, feeling happiness, and perceptions of self-efficacy (Holopainen et al., 2018; Jing et al., 2016; Martyr et al., 2018; O'Rourke, Duggleby, et al., 2015). Factors *negatively* associated with well-being and QoL is self-rated depression, sadness and anxiety, early onset of dementia, dependency, loneliness (Martyr et al., 2018), lack of participation in meaningful activities, and lack of attachment to the environment one lives in (O'Rourke, Duggleby, et al., 2015). The caregiver burden of professional (Holopainen et al., 2018) and family caregivers (Cheng, 2017), are directly associated with the self-rated QoL of the person living with dementia. When caregivers are overwhelmed, the person with dementia also report lowered well-being.

Sociodemographic factors have not been systematically associated with self-reported QoL, with one exception: living arrangement. People who live at home (Martyr et al., 2018) or in home-like environments in long-term care (Holopainen et al., 2018; Jing et al., 2016), rate their QoL better. For people living in long-term care, staff knowledge about dementia affects the quality of care, and hence, the self-rated QoL (Jing et al., 2016; Martyr et al., 2018).

When evaluated by *proxy-raters*, some additional aspects are negatively associated with well-being and QoL in people with dementia. These are functional dependency, dementia severity, and level of behavioral disturbances (Jing et al., 2016; Martyr et al., 2018). Relationship factors are essential in this equation as well. The family caregivers' self-rated QoL is associated with the self-rated QoL of the people living with dementia (Martyr et al., 2018). The coping strategies chosen by family caregivers predict the progression of dementia, with beneficial styles being problem-solving approaches and looking at the situation in a positive way (Tschanz et al., 2012). The relationship closeness between dyads also predicts slower cognitive decline in people with dementia, especially for spouses (Norton et al., 2009). The self-efficacy of the caregiver also positively influences the relationship (Cheng et al., 2013).

In sum, relationship factors, social factors, living at home, and experiencing meaningful enjoyable activities are core factors associated with self-reported well-being and QoL. Pinpointing which mechanisms and interventions that help people living with dementia to reach the best possible functioning is in line with the positive psychological perspective that is called for in several recent dementia initiatives (Gaugler et al., 2019; Lord et al., 2020; Reilly et al., 2020).

1.5 Interventions targeting well-being in dementia

Although several effective interventions indirectly improve well-being by targeting caregiver skills or environments only (Livingston et al., 2017), this thesis focuses on the interventions directly involving people living with dementia.

1.5.1 Pharmacological interventions targeting well-being

Certain pharmacological approaches may positively affect cognition, such as cholinesterase inhibitors for mild to moderately severe AD and DLB and memantine for some people living with severe AD (Livingston et al., 2017). Psychotropic medication targeting adverse symptoms, can at its best, ease the most severe BPSD, but the prescription is problematical because of harmful and highly frequent side effects (Livingston et al., 2017). The recommended approaches to support well-being in dementia are psychosocial and environmental interventions, and no pharmacological interventions have proven effective in *increasing* QoL and well-being (Abraha et al., 2017; Cooper et al., 2013; Livingston et al., 2017; McDermott et al., 2019).

1.5.2 Psychosocial interventions targeting well-being

Central psychosocial interventions developed to target well-being in dementia include cognitive stimulation therapy, reminiscence therapy, validation therapy, animal therapy, garden therapy, and creative therapies such as art therapy, dance therapy or music therapy, physical training interventions, and different approaches including sensory stimulation (Abraha et al., 2017; Kishita et al., 2020; McDermott et al., 2019). The biopsychosocial model shows that symptoms and needs in people with dementia are complex and individual, arising from numerous causes (Kales et al., 2015; Spector & Orrell, 2010). Thus, multidisciplinary, multimodal, and tailored approaches are required to be effective for people living with dementia on an individual level (Livingston et al., 2017). Interventions targeting adverse symptoms are important, but as underlined earlier, well-being is something more than the absence of negative symptoms (Martyr et al., 2018).

While the evidence of the different psychosocial interventions varies, some common elements seem beneficial for facilitating well-being. It is beneficial for interventions to support independence and agency and support functional ability (Beerens et al., 2016; Holopainen et al., 2018; O'Rourke, Duggleby, et al., 2015), include social interaction and facilitate connection (Livingston et al., 2017; Norton et al., 2009), deliver tailored, adjusted, and pleasant activities (Brod et al., 1999; Logsdon et al., 2002; O'Rourke, Duggleby, et al., 2015), and provide cognitive and sensory stimulation adapted to the optimal level of the person living with dementia (Dowson et al., 2019; Lobbia et al., 2019; Smith et al., 2004). One of the psychosocial approaches that involve these elements simultaneously is music therapy.

1.5.3 Therapeutic features of music

A growing evidence base supports that music therapy can contribute to neurorehabilitation in several conditions such as dementia, stroke, Parkinson’s disease, Multiple Sclerosis, epilepsy, and autism spectrum disorder (Brancatisano et al., 2020; Sihvonen et al., 2017; Särkämö, 2018). Neurorehabilitation refers to strategies of improvement, recovery, or compensation of impaired functions in the brain (Sihvonen et al., 2017). In their recently proposed Therapeutic Music Capacities

Model (TMCM), Brancatisano et al. (2020) describe the therapeutic factors of music in neurological disorders. The model describes in which contexts music may be therapeutic and outlines seven capacities of music. Music *engages* on a psychological and neurological level. Neurological engagement is supported by activating multiple brain regions and stimulating cognitive functions such as attention, memory, and learning. Music is *emotional*, mostly associated with positive affect. Music is *physical*, automatically encouraging the listener to move. Music is *synchronous* through its rhythmical or melodic elements. This aspect may be beneficial for speech in aphasia that is common in AD and VD. It can also benefit motoric and balance, often affected in AD, VD, and LBD. Music is *personal* and may facilitate positive self-perceptions that otherwise can be low because of the stigmatization or loss of role functions in people with dementia. Music is *social,* facilitating relationships, attachment, and nonverbal communication between people. Musical activities in groups may relieve social isolation, and music can bring people together in activities that can be managed despite cognitive impairment. Finally, music is *persuasive* because it may bring inspiration, optimism, and reinforce beliefs. This point is illustrated through the widespread use of music in religious contexts and marketing. These topics outlined by Brancatisano et al. (2020) are supported by research investigating neural mechanisms, physical, psychological, and social benefits of music and fit well with the tractable factors of the biopsychosocial model of dementia. An overview of the neural and physical mechanisms is outside the scope of this thesis, but several reviews have been published in the last five years (i. e., Brancatisano et al., 2020; Clements-Cortes & Bartel, 2018; Sihvonen et al., 2017; Särkämö, 2018). However, some findings from research on the psychological and social benefits of music therapy in dementia are central to review.

Evidence from meta-analyses of randomized controlled trials (RCTs) of dementia has demonstrated that music therapy has a beneficial effect for reducing depression and anxiety in dementia (Abraha et al., 2017; Kishita et al., 2020; van der Steen et al., 2018). The positive effect from music on depression and apathy in dementia may be related to its activation of systems of reward and arousal in the brain (Sihvonen et al., 2017; Särkämö, 2018), as well as the potential of music to easily activate positive emotions (Juslin, 2013). The positive effect on anxiety may be related to increased parasympathetic activation and inhibition of sympathetic activation, linked to lowered bodily activation and a decrease in stress hormones (Sihvonen et al., 2017).

Music therapy seems to have positive effects on overall cognition when the intervention is active (Fusar-Poli et al., 2018), although the evidence of an effect on cognitive functioning still is imprecise, conflicting (Fusar-Poli et al., 2018; van der Steen et al., 2018), and probably reflects a momentary improvement (Särkämö, 2018). It is suggested that the effect on cognition is related to accessing cognitive reserve, and the extensive activation of the brain network that can arise from musical stimulation (Sihvonen et al., 2017; Särkämö, 2018).

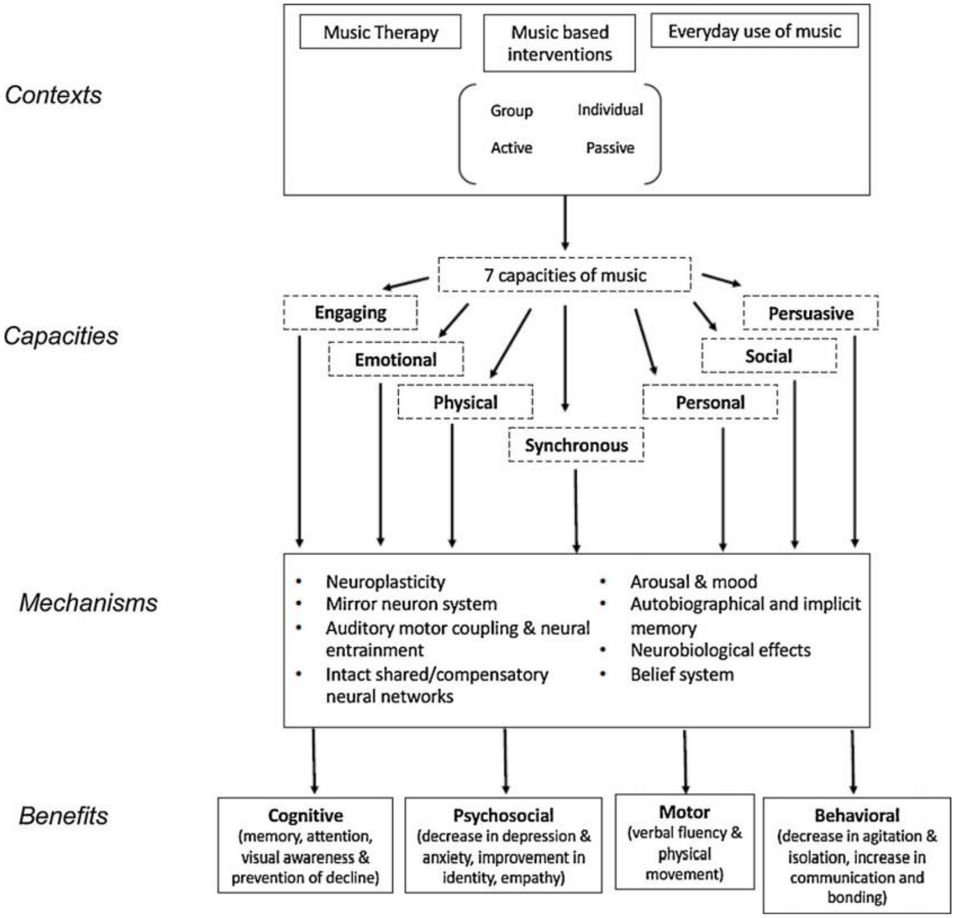
Music therapy appears to have a small positive impact on well-being and QoL in dementia when assessed with quantitative methods. However, this assumption is based on only a few RCTs which is, at present, evaluated as an imprecise conclusion that requires more trustworthy evidence (van der Steen et al., 2018). Qualitative music therapy studies reveal several benefits related to well-being, including pride, confidence, social connection, belonging, and overall positive emotions (Dowlen et al., 2018). Central mechanisms accounting for the well-being potential in music therapy are suggested. First, music seems to facilitate positive self-perceptions in dementia (Baird & Thompson, 2018). Additionally, familiar music triggers music-evoked autobiographical memories (called MEAMs) in people with AD, which are spontaneously and often emotionally pleasant memories from a person’s life (Cuddy et al., 2015). Musical memories seem to be especially well preserved in AD and bear the potential of reinforcing a positive identity (Cuddy et al., 2017; El Haj et al., 2013; El Haj et al., 2012). Creative skills and musical memory seem to be relatively well-preserved despite the dementia progression. The experience of mastery and confidence that may arise from being independent and feeling competent during musical activity is valuable (Baird & Thompson, 2018). Additionally, music therapy may be easily individually tailored, and the combination with nostalgic reminiscence elements seems to be of therapeutic value (Istvandity, 2017).

The abovementioned positive effects are mostly relevant for targeting the individual consequences of dementia. Furthermore, there is research supporting music therapy as having the potential to target some of the relational and interpersonal consequences of dementia. RCTs suggest music may positively affect social and empathetic behavior in dementia (Raglio et al., 2008; Särkämö, 2018; van der Steen et al., 2018). However, this conclusion is very uncertain (van der Steen et al., 2018). Single-case music interventions have found increased communication behavior (Schall et al., 2015), mutual engagement (Clair, 2002), and increased positive verbal and nonverbal communication (Engstrom et al., 2011) during tailored music interventions. Qualitative studies have described group singing to facilitate positive relationships with family caregivers (Clark et al., 2018; Mittelman & Papayannopoulou, 2018; Osman et al., 2016). Because BPSD are troublesome for persons living with dementia and their caregivers (Kales et al., 2015), the potential of music therapy to modify some of the many causes of BPSD is widely investigated. Music therapy may alleviate the general level of BPSD, but the effect on agitation and aggression is uncertain (Abraha et al., 2017; van der Steen et al., 2018).

In sum, looking at the individual and relational consequences of dementia previously reviewed, the therapeutic potential of music is significant. Therapeutic music interventions seem to access brain areas less impaired by the dementia, so cognitive, psychosocial, motor, and behavioral benefits may be within reach (Brancatisano et al., 2020; Sihvonen et al., 2017).

Still, several unresolved issues have been identified for music therapy assessment. First, most studies investigating music therapy in dementia assess group music interventions in long-term care facilities (Särkämö, 2018; van der Steen et al., 2018). The music interventions are often poorly described and lack individualization of the musical activity. Many studies have a small number of participants, do not investigate long-term effects, and does not properly distinguish music therapy led by a formally educated music therapist from music interventions provided by other professionals (McDermott et al., 2013; Sihvonen et al., 2017; Särkämö, 2018; van der Steen et al., 2018).

Figure 2 – The Therapeutic Music Capacities Model (TMCM)



Source: Brancatisano et al. (2020; p 602). Reprinted with permission from Elsevier®

Two other concerns regarding assessing music therapy interventions in dementia are also common for assessing psychosocial interventions for dementia in general. The first issue is the inconsistency of outcome measures, and the next issue is the lack of sensitivity in the outcome measures or study designs. These issues are elaborated on in the following section.

1.6 Issues when evaluating psychosocial interventions

The variety in outcome measures used in music therapy and psychosocial intervention research in dementia have challenged drawing clear conclusions in meta-analyses or systematic reviews (Abraha et al., 2017; Kishita et al., 2020). Inconsistency in conceptualizations, operationalizations, and lack of data on measurement properties may complicate the interpretation and trustworthiness of research findings. Recently, several projects such as the COMET initiative (www.comet-initiative.org) and the ICHOM initiative (www.connect.ichom.org/standard-sets/dementia/) have focused on developing valid, comprehensive, and relevant core outcome sets for dementia research and care. Well-being and QoL have been identified by both initiatives as one of several important outcomes. The vast number of measurement instruments of QoL and well-being are unfortunately not well supported concerning validity and reliability (Algar et al., 2016; Bowling et al., 2015; Hughes et al., 2019; Madsø & Nordhus, 2021; Missotten et al., 2016; Santana-Berlanga et al., 2020). The recent systematization of relevant and consistent outcome measures for psychosocial intervention research has focused on self-reported outcomes within positive psychology and social health (Harding et al., 2019; Reilly et al., 2020) and self-reported momentary well-being (Clarke et al., 2020). Less attention has been given to *observational measures* of well-being with a momentary focus.

Furthermore, benefits of music therapy and other psychosocial interventions are mostly assessed at the end of treatment, and less is known about long-term benefits and mechanisms of change (McDermott et al., 2019; van der Steen et al., 2018). Both qualitative and quantitative inquiries seem to agree that the strongest benefits from music therapy are “in the moment” (Dowlen et al., 2018; Särkämö, 2018), but many intervention studies focus on benefits measured by aggregated global recall instruments. Because of the progressive nature of dementia, using global scores at the end of treatment may mask the momentary benefits from music therapy and other relevant psychosocial interventions for people living with dementia (Kishita et al., 2020). As qualitative inquiries seem to reveal benefits that the quantitative studies have yet to support, strengthening the sensitivity of the study design and measurement instruments is more likely to filter out any masking effect of confounding variables. Assessing clinically significant changes *as they unfold* seems advantageous. This approach is also recommended when assessing well-being (Clarke et al., 2020; Diener, 2006). Momentary measurement may be particularly useful in dementia research and assessment. The benefits of momentary assessment and momentary observational methods are elucidated in the next section.

1.6.1 Ecological momentary assessment

Ecological momentary assessment (EMA) is defined as *'methods using repeated collection of real-time data on subjects' behavior and experience in their natural environments'* (p 3, Shiffman et al., 2008). It includes data from event sampling methods, diaries, natural observation, or video-recordings. Shiffman et al. (2008) describe three general advantages of EMA. First, it is not hampered by memory bias, which may lower the reliability of rating-scales globally rating wider time-periods such as "the last week." Second, it has high ecological validity because it is conducted in real-life situations. Third, it may capture clinically relevant microprocesses over short time lags. Proxy-ratings for QoL or well-being correspond poorly to self-ratings (Ferring & Boll, 2010; Schulz et al., 2013), and momentary assessments may enable people with dementia to self-report reliably even with increasing dementia severity because it is not dependent on long-term memory (Clarke et al., 2020). The microprocesses that can be assessed in EMA approaches may be particularly advantageous for assessing positive changes in dementia. The biopsychosocial model shows that some causes of ill-being in dementia may be more chronic (Kales et al., 2015; Spector & Orrell, 2010). However, because well-being can also be independent of dementia severity and ill-being, momentary well-being may coexist with chronic or fluctuating ill-being. Momentary measures can detect changes that may otherwise be masked by the fluctuating symptoms, tractable psychosocial or biological aspects, and the progressive nature of dementia - if evaluated over long-term intervals (Clarke et al., 2020; Kales et al., 2015; Kolanowski et al., 2017; Spector & Orrell, 2010).

While self-reporting is most common in the general population when assessing well-being, recommendations for people with dementia are to facilitate self-reporting when feasible (Clarke et al., 2020). Using neutral observers to assess well-being when self-report is not feasible is recommended to avoid the pitfalls of proxy-ratings mentioned above (Ferring & Boll, 2010). The three advantages of EMA described in the former paragraph are also valid for momentary observational measures.

1.6.2 Central characteristics of observational approaches

Direct systematic observation has long traditions in empirical psychological research, resulting in qualitative narratives and quantitative coding assigning numbers to measure behavior (Bakeman & Quera, 2011). Observational methods are particularly suited for four contexts: first, when the participants cannot readily share their thoughts or respond to questionnaires; next, when the behavior of interest is nonverbal; third, when the behavior of interest is occurring naturally and spontaneously; fourth, when we are trying to understand how behaviors unfold over time, the mechanisms of change, or the contingency (Bakeman & Quera, 2011). These contexts often apply to dementia research, and systematic observation is commonly employed (Algar et al., 2016; Curyto et al., 2008). Table 2 displays the central characteristic definitions of observational measures that are relevant to the applicability of the instruments for different objectives.

Table 2 – The defining characteristics of observational instruments

Characteristic	Description	
Type	Observational rating scale Ethogram Coding scheme	
Granularity	From fine-grained moment to moment changes to broad grained overall summary states	
Concreteness	From physically observable motions to more social inferred states	
Complexity	From global levels of the constructs to more complex behavioral structures	
Measurement level	Nominal (i.e., mutually exclusive categories, dichotomous present/absent), ordinal (i.e., rating scales, Likert scales, the stronger intensity with an increasing number) interval or ratio (i.e., the proportion of time). This approach defines your data-analytic plan – what research questions you can answer.	
Recording level	In vivo (as behavior unfolds) or collected by digital device (audio or videotape)	
Primary recording unit	Events – continuous recording Intervals – interval recording(time-sampling)	Continuous untimed (multi) event recording Continuous timed-event recording
		Partial-interval sampling (zero-one sampling) Momentary sampling Whole-interval sampling
Development of observational instruments	Top-down process: Theory-driven, based on literature reviews and knowledge about a phenomenon. Based on consensus in expert groups, Delphi-technique, or survey-feedback from users. Adaptations of other coding schemes. Bottom-up process: Qualitatively derived categories through observations, systematic development of ethograms, qualitative interviews with experts or users. Mixed methods	

Note: This table is based on Ayres and Gast (2009); Bakeman and Quera (2011, 2012)

1.7 Summary

To sum up, dementia is a highly prevalent condition world-wide, with vast individual, relational and societal consequences. Optimizing the possibility of living well with dementia requires focusing on which tractable biological and psychosocial factors to intervene on, which interventions to use, and when to use them. Based on the initiatives that call for positive psychological perspectives within dementia research and care (Dröes et al., 2016; Gaugler et al., 2019; Lord et al., 2020; Reilly et al., 2020), this thesis seeks to contribute to the knowledge base on positive and resource-oriented approaches for people living with dementia.

Emotional, social, and psychological well-being are central aspects that people with dementia define as important outcomes for psychosocial interventions. Reviewing observational instruments used to assess well-being is needed. It may provide updated knowledge for researchers and practitioners about which measures work best in which contexts and facilitate the use of sensitive measures assessing change as it unfolds.

Regarding interventions, music therapy is a promising approach for people living with dementia. However, investigations of music therapy for the home-dwelling population are missing. This population makes up the largest share of the people living with dementia today. Additionally, there is a call for research assessing the positive aspects of music therapy in dementia, providing adequate descriptions of the interventions, and delivering individualized approaches.

2. Aims

The overall aim of this thesis was to add to the scientific knowledge about well-being for people living with dementia: first, providing a comprehensive systematic overview of observational approaches for assessing momentary well-being in people living with dementia; next, investigating the potential of music therapy to facilitate clinically relevant positive changes in well-being and sociable interaction for people living with dementia and their significant others; finally, developing and field-testing an observational instrument suitable for comparing music therapy and other psychosocial interventions.

2.1.1 Research questions and aims for Paper 1

The aim of Paper 1 was to identify the relevant observational instruments assessing momentary well-being in people living with dementia and assess study-specific and instrument-specific measurement properties against pre-defined quality criteria. Next, the aim was to grade the trustworthiness of the available evidence for each instrument. Lastly, the paper aimed to evaluate the feasibility and interpretability of the instruments to inform implementation in research and care.

2.1.2 Research questions and aims for Paper 2

The aim for Paper 2 was to compare music therapy with regular sociable interactions for home-dwelling people living with dementia together with a family caregiver. The primary hypotheses were that a) the level of observed well-being was higher during music therapy compared to regular social interaction; b) the level of observed sociable behavior towards significant others was higher during music therapy compared to regular social interaction. The secondary hypotheses were that a) the self-rated positive emotions would increase, and self-rated negative emotions would decrease, from pre- to post music therapy sessions; b) the neuropsychiatric symptoms would be stable from pre- to post 10-week intervention; c) the self-reported long-term well-being would increase from pre- to post 10-week intervention; d) caregiver burden would decrease from pre- to post 10-week intervention.

Furthermore, Paper 2 examined the utility of single-case designs to investigate clinically relevant short-term effects of music therapy in people living with dementia.

2.1.3 Research questions and aims for Paper 3

The aim of Paper 3 was to develop, and field-test, an observational instrument and its potential feasibility for comparing observed well-being in people living with dementia during music therapy and social interaction.

The objective was to establish the content validity of the instrument through an iterative process of anchoring the instrument in theory and assessment of other observational well-being instruments, and conducting qualitative fieldwork to develop quantifiable items. Next, the aim was to assess relevant psychometric properties of the instrument in a field-test; inter-rater reliability and agreement, construct validity, and responsiveness. The assessment of the instruments' content validity in the eyes of experts was the next aim, assessed in focus groups. A new revision was then field-tested to establish feasibility and assess intra-rater reliability.

3. Methods

Mixed methods were employed in this thesis, which included using quantitative and qualitative data collection, analytic approaches, and quality standards (Johnson et al., 2007). The choice of methods was pragmatic, based on which approach provided the best options to answer the research questions and aims (Dures et al., 2010; Johnson & Onwuegbuzie, 2004; Morse, 2010). Central issues regarding the mixed methods approach are elaborated in the first part of this methods chapter, followed by a description and justification of the specific methods and procedures in the three papers.

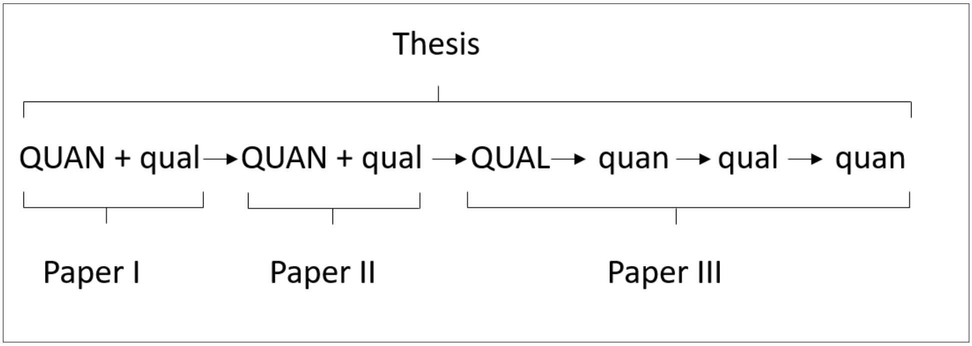
3.1 Mixed methods - quality standards

Quantitative and qualitative methods have divergent and sometimes incompatible quality standards (Wiggins, 2011). Three main steps were taken in this thesis to conduct a trustworthy and valid mixed methods design. The first step included transparency about the sequencing, priority, and purpose of the distinct methods and their mixing as described in section 3.1.1 (Creswell, 2015; Mark, 2015; Morse, 2010). The next step was selecting separate quality criteria for each qualitative and quantitative phase, followed by specifying quality criteria and sources of bias specific to the mixed methods approach of the thesis, as presented in section 3.1.2 (Collins, 2015). Utilizing mixed methods can add scientific rigor when the strengths in one approach compensates for the limitations of another approach. However, because the mixed methods in this thesis endorse an iterative approach where inductive and deductive logics are used, it can lead to scientific threats to validity (Creswell, 2015; Johnson & Onwuegbuzie, 2004; Teddlie, 2010). Thus, a third step was to describe and justify the different methods and levels of data integration, enabling the reader to evaluate the legitimacy of the approach, as described in section 3.1.3 (Fetters et al., 2013; Mark, 2015).

3.1.1 Sequencing, priority, and purpose

This thesis follows a multi-staged mixed methods framework where Paper 1 and 2 analyze data simultaneously or parallel in a *convergent design.* Paper 3 utilizes an *exploratory sequential* data collection approach where each step *builds* on the analysis and results from the former (Fetters et al., 2013). The overview of the mixed methods sequence and priority in this thesis and the three papers, are based on Morse (2010) and is outlined in Figure 3.

Figure 3 – Overview of the mixed methods approach



The '+' in Figure 3 describes when quantitative and qualitative methods or criteria are used in parallel, and the arrow describes when the method is following the other sequentially. The capital letters define which methodological approach dominates in the paper, where *QUAN* describes explanatory approaches and *QUAL* describes exploratory approaches. For Paper 1, the main approach was evaluating well-being measures suitable for quantitative observational investigations of changes in well-being for people living with dementia with *QUANT*itative quality criteria (Mokkink et al., 2018). However, the *qual*itative aspect simultaneously assessed was the evaluation of content validity, where the gold-standard is qualitative fieldwork or conducting qualitative focus groups or individual interviews (Chorney et al., 2015; Terwee et al., 2018). The purpose of mixing these quantitative and qualitative evaluations was to assess each measurement property with the method best suited to investigate the respective properties, with a total integrated evaluation of the instruments displayed together.

For Paper 2, the focus was on *QUANT*itatively assessing music therapy from systematic observations, as well as with other quantitative assessment scales. However, the assessment of well-being was based on a *qual*itative inquiry of the observable components of well-being. These qualitative themes were assigned numbers through coding. The purpose of mixing methods was to make sure the well-being conceptualization was grounded in well-being as a phenomenon during music therapy in dementia and make the data available for statistical assessment.

For Paper 3, the dominating approach was inductive and aiming at discovering patterns (Johnson & Onwuegbuzie, 2004). The development and refinement of the well-being measure were based on *QUAL*itative fieldwork. In contrast, the process of investigating measurement properties was based on the deductive testing of hypotheses with *quant*itative methods. A supplemental *qual*itative component elaborating content validity and feasibility of the instrument in focus groups was building on the former steps, followed by a *quant*itative field-test. The purpose of integrating the data from the different phases through the *building* (Fetters et al., 2013) was to elaborate and establish the content validity of the instrument. This iterative process is intended to optimize the measurement properties and feasibility of the instrument (Bakeman & Quera, 2011).

3.1.2 Quality criteria for the separate and mixed methods

The nomenclature of scientific rigor in the different methodological traditions differ; internal validity in quantitative research corresponds to the trustworthiness of qualitative research and legitimation in mixed methods, while external validity in quantitative research corresponds to credibility in qualitative research and inference quality in mixed methods (Collins, 2015). It was decided to choose separate criteria to assess the quality of research in each phase. This process involved evaluating the scientific rigor associated with the specific methodological approach, then evaluating validity threats in the mixing of methods, analyses, or results (Collins, 2015; Creswell, 2015; Fetters et al., 2013).

In line with Collins (2015), the separate quality criteria for the different methods are outlined. For Paper 1, these quality criteria are based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Liberati et al., 2009) and the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN; Mokkink et al., 2020; Mokkink et al., 2017; Prinsen et al., 2018; Terwee et al., 2018). The quality criteria for Paper 2 are mainly built on the Single-Case Reporting Guideline in Behavioral Interventions (SCRIBE; Tate et al., 2016). The quality criteria for Paper 3 is based on COSMIN for the quantitative and qualitative part, with the latter supplemented with standards of rigor of qualitative research as suggested in the National Institute for Health and Care Excellence guidelines (NICE; 2015). These guidelines are elaborated under each methodic description in the following sections.

Thre relevant threats to validity in the mixed methods are identified. The first is the quantitizing of the qualitative themes in Paper 2 and 3. The second is relying on induction in the qualitative fieldwork followed by deduction in the quantitative assessment in Paper 2 and 3. The last is the sampling procedures in the qualitative and quantitative assessment of Paper 2 and 3 (Collins, 2015; Creswell, 2015; Onwuegbuzie & Johnson, 2006).

3.1.3 Justification of methods and data integration

The third step in this thesis to conduct a trustworthy mixed methods design is to justify each method and legitimate the inferences derived from the data integration in the mixed methods approach (Collins, 2015). Each methodical choice and the specific data integration are elaborated under the methodical description for each paper individually in the remainder of the methods chapter in this thesis. The legitimation of the inferences from the results of the mixed methods are elaborated on in the discussion.

3.2 Methods and procedures for Paper 1

In the first paper, available observational instruments assessing well-being in people living with dementia were reviewed utilizing the COSMIN approach for conducting systematic reviews of health-related outcome instruments (Prinsen et al., 2018).

3.2.1 Systematic literature search

A protocol describing the aims, eligibility criteria, search strategy, and the data-extraction plan was pre-registered in the International Prospective Register of Systematic Reviews (PROSPERO, ID: 17160). Registration records during 2020 were published without assessment of the PROSPERO editorial team because of the prioritization of COVID-19 registrations.

Eligibility criteria

The studies included for review described observational measures assessing well-being with a momentary timeframe (from seconds to hours). The instrument should be used for direct or video-recorded observations and assess well-being before and after or during any psychosocial intervention. It should be feasible for independent observers instead of family and other professional caregivers with an established relationship with the person living with dementia. The operationalizations should focus on well-being, defined as behavioral expressions of satisfaction, or positive and negative emotions, in line with the definition from Diener (1984). At a minimum, at least one psychometric property should be assessed. Assessed measurement properties were defined after the COSMIN-definition criteria (Prinsen et al., 2018). Instruments should be dementia-specific or tested in a sample of people with dementia. Instruments primarily focusing on ill-being or physiological measures of well-being were excluded. Only English peer-reviewed journal articles were assessed.

Search strategy

The search string was developed in collaboration with the main supervisor (IHN), and I consulted the research literature to identify proper search words and synonyms. To ensure all publications regarding measurement properties were detected in our search, a published search filter for this purpose was utilized (Terwee et al., 2009). Additionally, an experienced librarian guided the search procedures to ensure they were adapted to the specific databases and included the correct Boolean operators and truncations. The searches were conducted in PsycINFO, Medline, Embase, Web of Science, CINAHL, and ProQuest (sections Psychology and Nursing and Allied Health). The search string combined different words reflecting "well-being" AND "dementia" AND "observation" OR "measurement" OR "psychometric properties." A hand search of reference lists, relevant reviews, and forward chaining citations of the included publications was conducted to detect other relevant articles describing relevant instruments.

Study selection

The search was conducted twice, on the 21st of April 2020 and on the 6th of April 2021. The selection of studies involved several steps. First, all records were imported to Endnote ® and deduplicated. Next, titles and abstracts were screened against eligibility criteria, and all records including people with dementia, observation, well-being, or reviews of instruments, were kept for the next stage. Reviews and other relevant records were hand-searched. Then, the remaining records and the results from hand searches were imported to Rayyan, where the abstracts were blind screened by me and the main supervisor (IHN). The blinds were removed, and the remaining records were full text assessed for eligibility by me and IHN in collaboration to identify the instruments and records included for evaluation.

3.2.2 Data extraction

The extraction-procedure was conducted following the COSMIN Risk of Bias (RoB) checklist (Mokkink et al., 2017) and adaptations to extract relevant information about observational measures (Bakeman & Quera, 2012). Data were extracted into five tables containing a) characteristics of the measure, b) characteristics of the included study populations, c) interpretability of the instrument, including distribution, skewness, statistical sensitivity, and clinical utility, d) feasibility of the instrument, e) methods and results of the publications reporting the following measurement properties: content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, construct validity through hypothesis testing and responsiveness (de Vet et al., 2011; Mokkink et al., 2017; Prinsen et al., 2018; Terwee et al., 2018). I conducted data extraction, which was reviewed in consensus-meetings with the main supervisor (IHN). Three from the team of authors (KGM, EFG and IHN) synthesized the extracted data in to the three results tables in Paper 2.

Risk of Bias assessment at the *study level*

The extracted data was assessed regarding content validity on study level with an adapted RoB-approach. The COSMIN assessment criteria were designed for patient-rated outcome measures (PROMS; Terwee et al., 2018), not observational measures. Field-testing is particularly important in developing coding schemes, as theoretically important domains may work poorly for real-life observations (Bakeman & Quera, 2011). The COSMIN guidelines for content validity were adapted for observational instruments through an extensive and theoretically based process (Bakeman & Quera, 2011, 2012; Chorney et al., 2015). An overview of the adapted RoB-checklist to assess development studies and content validity is provided in the Appendix A. RoB of the remaining psychometric properties on study level were evaluated with the COSMIN-approach (Mokkink et al., 2020; Mokkink et al., 2017; Prinsen et al., 2018). These included structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, construct validity through hypothesis testing and responsiveness, and the specific RoB-items on study level are available in the COSMIN user manual (Mokkink et al., 2018).

Evaluation of evidence against quality criteria at the *instrument level*

Like the RoB-ratings on the study level, the quality criteria on instrument level had to be adapted for observational instruments. The main changes were made to the assessment of content validity. We decided that a lack of pilot-testing of the final version of the instrument would lead to an “inadequate” rating of total content validity. Additionally, *at least* two of the following approaches were required in the development of the instrument: a theoretical approach with a literature review, adaptations of other coding schemes, qualitative fieldwork and development of a coding scheme or ethogram, quantitative survey or qualitative interviews and focus groups including target group (experts from all relevant disciplines and/or patients and family care givers; Terwee et al., 2018; Bakeman & Quera, 2011; Bakeman & Quera, 2012; Chorney et al., 2015; Perugia et al., 2018). It was decided to use a cutoff of at least ≥.3 for correlations to be judged as supportive evidence for the construct validity and responsiveness of the instruments, based on the generic hypotheses suggested in Prinsen et al. (2018). All changes made to the original quality criteria are specified in Table 2 of Paper 1.

Rating the trustworthiness of the evidence at the instrument level

The COSMIN-framework for systematic reviews utilizes an adapted Grading of Recommendations Assessment, Development and Evaluation (GRADE)-approach for evaluating the trustworthiness of the available evidence of each outcome measure. The four factors constituting the GRADE-assessment are 1) the level of RoB ("very good", "adequate", "doubtful" and "inadequate"), 2) inconsistency, 3) imprecision, and 4) indirectness (Prinsen et al., 2018). For each factor, the level of evidence can be downgraded according to the evidence quality, leading to a "high," "moderate," "low," or "very low" rating.

Evaluation of feasibility and interpretability at the instrument level

Instrument-specific details about feasibility and interpretability were extracted to inform clinicians and researchers about the total clinical utility of the instruments in different contexts. The key information was the instruments' granularity, concreteness, measurement level and primary recording unit (Ayres & Gast, 2009; Bakeman & Quera, 2012), as defined in Table 2 in the introduction. A hand search was conducted to assess the different instruments used in interventional research, gain and extract information about skewness, and ability to detect statistically or clinically relevant changes. These were added to the results to inform the interpretability, and aid implementation of the recommended instruments.

3.2.3 Statistical assessment

Except for calculating the percentage of supported hypotheses investigating construct validity for each instrument, no statistical pooling of results across the included studies was conducted. When several sources were reporting on the same measurement property for the same instrument, the evaluation against quality criteria was based on the scores' total range unless otherwise specified.

3.3 Methods and procedures for Paper 2

The second paper tested a music intervention for people living with dementia using a replicated single-case design, momentary pre-post measures, and pre-post intervention measures.

3.3.1 Sample and procedures

Participants were mostly recruited from NKS Olaviken Gerontopsychiatric Hospital in Bergen, Norway, or through an ad in the local newspaper, and brochures available in relevant arenas. Interested participants had to be referred to the outpatient clinic to participate. Inclusion criteria were 1) diagnosis of AD, VD, DLB, dementia with mixed ethology, or Parkinson's Disease Dementia according to the criteria of ICD-10 (WHO, 2016); 2) mild to moderately severe dementia (0.5 - 2 on the Clinical Dementia Rating (CDR); Hughes et al., 1982); 3) a caregiver committing to contribute as a collateral therapist; 4) stable psychotropic medication at least two weeks before pre-assessment and during the intervention-period; and 5) living at home or in a home-like assisted living environment. Exclusion criteria were 1) FTD; 2) severe aphasia; 3) severe psychotic symptoms or high risk of suicide; 4) permanent or temporary living arrangement in long-term care.

Initially, a plan to recruiting 20 participants was made, based on the single-case research standards of Kratochwill et al. (2012). After performing a quality testing of the intervention for one participant and caregiver in 2016, 13 participant-dyads were enrolled from January 2017 to December 2019. Three participants dropped out, and one participant later re-entered the study. Of the 11 participants, 63% were women.

They were aged 71-88 years (m = 79.82, SD = 5.27). Collateral therapists were three children and eight partners or spouses. One participant dropped out after six sessions, but all available data were included in the single-case analyses.

3.3.2 Intervention

Previous studies indicate a dose-response relationship predicting the general effect of music therapy, where moderate effects are achieved after 10 – 24 sessions, and large effects are achieved after 16 – 51 sessions (Gold et al., 2009). Thus, we planned for 20 sessions during the intervention, 10 with the music therapist and 10 with the collateral therapist. This approach increased the total session number and allowed the dyad to practice musical activities outside the therapy setting.

The music therapist assessed the musical history of the participants throughout their lifespan to tailor the music therapy to the participant. The music therapy and musical sessions with the collaterals were built around this assessment, with individual goals set by the participants. Musical elements of the interventions are provided in Table 3.

Table 3: Elements in the music intervention

Element	Description
Singing	Actively using the voice with and without word pronunciation
Playing music	Interactions with known music or improvisation together with music therapist (and collateral)
Listening to recorded music	Active/conscious listening
Listening to live music	Active/conscious listening
Dancing	The (rhythmical) movement to music while sitting, swaying, or dancing together with collateral or alone
Reminiscence	Conversations about topics before, during or after musical activities. Topics, opinions, or wishes for a particular song, describing memories that come to mind.
Exercise to music	The rhythmical movement to music with exercises tailored to the individual's wishes or needs
Relaxation	Breathing or relaxing exercises together with music

The therapeutic approach of the intervention was based on a resource-oriented music therapy treatment manual, where the central focus is to facilitate the strengths, potentials, competence, and goals of the client. Music therapy is based on the preferences of the client and is used to foster positive emotions, emotion regulation, social relationships, and positive self-perception (Rolvsjord et al., 2005).

3.3.3 Measurements and procedures

Measurements

The study design was set up to measure changes on three levels. 1) The single-case design compared a five-minute baseline phase with a 30-minute music therapy phase for three sessions per dyad (1st, 5th, and 10th). The constructs observed were well-being and sociable interaction towards significant others. 2) Pre- and post-session measures compared levels of self-reported emotions immediately before and after the music therapy session and the collateral-led sessions. 3) Pre- and post-intervention measures compared levels of self-reported QoL and caregiver-reported neuropsychiatric symptoms and caregiver burden. The measurement instruments are presented in Table 4.

Table 4: Measurement instruments

Construct	Instrument	Reference	Scoring
Well-being	Observable Well-being in Living with Dementia Scale (OWLS, version 1.0)	Madsø et al (2021, manuscript under review)	Ranges from 0-10. A score >2 indicates well-being
Sociable interaction	Verbal and Nonverbal Sociable Interaction Scale- Care Receiver (VNVIS-CR)	(Williams et al., 2017)	Twenty-six items where a ratio of sociable and non-sociable items is calculated. Ranges from 0-13. Two subscales: Nonverbal sociable interaction and Verbal Sociable interaction
Emotions	Visual Analogue Mood Scale (VAMS)	(Stern et al., 1997)	Eight emotions are self-reported, ranging from 0-100mm on a line
QoL	Quality of Life in Alzheimer's Disease (QoL-AD)	(Logsdon et al., 1999)	Thirteen domains are self-reported. Ranges from 13 – 52.
Neuropsychiatric symptoms	Neuropsychiatric Inventory-Questionnaire (NPI-Q)	(Kaufer et al., 2000)	Twelve domains describing presence and level of different symptoms are scored, Ranges from 0 – 36
Caregiver burden	Relative Stress Scale (RSS)	(Greene et al., 1982)	Fifteen domains are scored, ranging from 0 – 60
Dementia severity	Clinical Dementia Rating (DCR)	(Hughes et al., 1982)	0.5 = very mild dementia, 1 = mild dementia, 2 = moderately severe dementia, 3 = very severe dementia
Cognitive impairment	Mini-Mental State Examination Norwegian Revision (MMSE-NR)	(Strobel & Engedal, 2008)	Range 0-30. Scores <24 indicate cognitive impairment

Single-case design

Single-case designs are conceived of as advantageous for investigating change following interventions within subjects, where the person serves as their own control (Manolov et al., 2014; Manolov & Moeyaert, 2017). The design is sensitive to detect changes, has high ecological validity, and is well suited for small samples and populations with dementia (Steingrimsdottir & Arntzen, 2015). In addition, synthesizing single-case research may help us better understand the variation in the effectiveness, the predictors of effectiveness of interventions, generate hypotheses, and point to areas where further research is needed (Pustejovsky, 2018).

The single-case design of Paper 2 was built upon an earlier music-therapy investigation using single-case time-series analysis (Schall et al., 2015). The single-case design compared the baseline (A) of regular sociable interaction between the dyad and the music therapist and the intervention (B) with music therapy with the dyad and music therapist. This bi-phasic design (Tate et al., 2016) was replicated for each participant three times. The AB comparison was later replicated, giving 32 sessions, nested within 11 cases.

Three video-graphed sessions per participant were coded with OWLS and VNVIS-CR using Noldus Observer XT (Noldus Information Technology, 2015) by three trained coders. Inter-rater reliability was assessed with Cohen’s Kappa, and observer agreement was assessed as a percentage of agreement for 20% of the video-material. This analysis was provided by Noldus Observer XT (Noldus Information Technology, 2015).

3.3.4 Statistical analyses of Paper 2

The single-case calculator of Pustejovsky and Swan (2018; version 5.0.) was used to calculate the effect size Log Response Ratio (LRR). The statistical analyses were conducted in R (R Core Team, 2020) and R studio (RStudio Team, 2020).

Single-case analysis

Several parameters are relevant for assessing intervention effects of single-case design, such as within-phase variability, change in level, presence of baseline and intervention trend, autocorrelation, the immediacy of effect, overlap between phases, outliers, and the observed compared to the expected pattern (Kratochwill et al., 2012; Manolov & Solanas, 2013). Several effect sizes have been developed for single-case analysis, each with its strengths and limitations (Manolov et al., 2014; Parker et al., 2011; Vannest & Ninci, 2015). However, correcting for parameters not present or not considering important present parameters is a threat to the validity of the analysis. Thus, the analysis should start with a visual evaluation of the data to decide which visual, nonoverlap or statistical method to apply to account for present parameters (Parker et al., 2011).

The visual analysis of the plotted data focused on the presence of trends in the baseline and intervention phase, the immediacy of change, and the variability of the scores. The presence of baseline and intervention trend was rare. Change was mostly immediate and indicated a presence of change in level and not slope. The large variability of the scores suggested using a statistical rather than a visual analytic approach (Manolov & Vannest, 2019; Parker et al., 2011; Vannest & Ninci, 2015). Several potential effect size calculations are limited by procedural sensitivities (Pustejovsky, 2019). The effect sizes Nonoverlap of All Pairs (Parker & Vannest, 2009), Baseline corrected Tau (Tarlow, 2016), and Percentage Exceeding Median (Ma, 2006) were evaluated as not applicable. Thus, using change in level as the main parameter seemed most appropriate.

The effect size Log Response Ratios (LRR) were evaluated as most feasible; it compares two mean levels, is not sensitive to sample size or length of observations, is scale-invariant, and the magnitude can be converted to the percentage of change to ease interpretation (Pustejovsky, 2015, 2018; Pustejovsky et al., 2019). The LRR-model comes with two central assumptions. First, the data must be stable within the phase (no autocorrelation), and second, observations should be independent (Pustejovsky, 2018). Furthermore, the individual variance estimates of LRR are biased for small samples (Pustejovsky, 2018). These assumptions are often violated in single-case research, as autocorrelation is common, observations are rarely independent, and small samples are a rule of thumb. Meanwhile, the major advantage of the LRR is the option of synthesizing several single-case effect sizes through meta-analysis. This approach provides four strengths. First, it increases the generalizability of the results. Second, it captures the overall magnitude of any treatment effect. Third, it describes the consistency of the findings through assessing heterogeneity across cases. Fourth, it allows for hierarchical analyses where robust variance estimation corrects for small sample bias and autocorrelation (Pustejovsky, 2018). Adding sensitivity analysis enables the investigation of whether the dependency of the observations influences the stability of the results (Fisher & Tipton, 2017; Tipton & Pustejovsky, 2015). In total, these steps bypass the problems caused by violating the model assumptions.

The AB comparison of OWLS and VNVIS-CR per session was analyzed with the LRR-calculator (Pustejovsky & Swan, 2018). Next, a cluster variance estimation with small-sample correction was conducted (Hedges, 2019; Tipton & Pustejovsky, 2015), including all the individual music therapy sessions. Finally, meta-regression analysis was conducted to investigate sources of heterogeneity, including the hypothesized increasing effect over time (session one, five, and ten), dementia severity (very mild, mild, or moderately severe), and depression (mild or moderate depression). As the AB-comparisons were clustered within 11 cases, sensitivity analyses were conducted to investigate the stability of the results of the former analyses when allowing for different correlations between the observations (Fisher & Tipton, 2017; Tipton & Pustejovsky, 2015).

Exploring behavioral content at item-level

We explored the relative frequencies of the items of OWLS and VNVIS-CR to assess the differences and similarities in behavioral content in the two phases. This number was calculated as the total number of occurrences divided by the total number of observations and was reported in %.

Analysis of pre-post measures

The analysis of the self-reported emotion from pre- to post-sessions was initially planned to be assessed using hierarchical linear modeling (Woltman et al., 2012), allowing for assessing changes on groups, individuals, and over time. Because the data could not meet assumptions of normal distribution, statistical analysis of the pre-post session and intervention measures instead utilized the non-parametric Wilcoxon signed-rank test with continuity correction (R Core Team, 2020).

3.3.5 Ethical considerations

The study protocol was registered before recruiting participants (www.clinicaltrials.gov, ID NCT03011723), and recorded at the System for Risk and compliance at the University of Bergen (RETTE, ID: R639). The Regional Committees for Medical and Health Research Ethics (REK) in Norway (2016/1374) approved the clinical study. Informed consent is required to participate in health research regulated by the Health Research Act (Helseforskningsloven, 2009) and the Helsinki declaration (World Medical Association, 2013). In dementia, consideration about the *capacity* to consent is of particular importance. However, at the same time, protection and safeguarding should not exclude people wanting to contribute to research from participating (Reid et al., 2018). Thus, special considerations were given to assess and obtain informed consent from the participants, but both dyad members gave consent on their own behalf when included.

The procedures of consent were a stepwise and continuous process, individualized to the person by considering the participants' personal values (Moye et al., 2007). First, people eligible for participation were allowed to learn about the project in writing and through an oral presentation, provided with a copy of the Participation Information and Consent-form. They independently registered their interest at the outpatient clinic. All written information was authored in an adjusted and straightforward language. Participants chose a designated contact person with whom the home visit was scheduled after registration of interest. In the home visit, the music therapist and a clinical psychologist gave information about the study, legal consent, and the right to withdraw. The persons with dementia and their close caregiver decided if they wanted to participate. If consent was signed, the assessment session was initiated on the same visit. The possibility to discuss what participating meant, including individual risks and gains that participating could give, was focused upon when meeting our potential participants. Throughout the intervention-phase, consent was viewed as an ongoing process. The music therapists and I discussed signs of withdrawal of consent throughout the participation. We would explicitly remind the participants of the research project during the intervention when needed. The principles for assessing consent were based on Dewing (2007), McKeown et al. (2010), Moye et al. (2007), and Warner et al. (2008).

3.4 Methods and procedures for Paper 3

The third paper reported the development, field-testing, and psychometric properties of the instrument Observable Well-being in Living with Dementia-Scale (OWLS).

3.4.1 Review of literature and existing instruments

The development of coding schemes fit for observational assessment is an iterative process based on the underlying theoretical understanding of what is important to measure (Bakeman & Quera, 2012). Thus, a theoretical overview of conceptual models of generic and dementia-specific models of well-being guided our initial development phase. The research questions in the music therapy intervention study guided the choice of measurement instruments. After conducting a literature review of available observational instruments (Paper 1), where no instrument fit our purpose, the decision to develop a new instrument was made.

3.4.2 Developing the initial items

The well-being theory and instruments found in the literature search were the starting point for detecting possibly relevant items. The qualitative fieldwork was conducted, where I repeatedly watched 10-minute video segments from the intervention study to detect significant sections of the video material (Heath et al., 2010). The sampling of sections was purposive (Malterud, 2001), including sections from several participants until the themes and descriptions were appraised as saturated (Braun & Clarke, 2006). The sampled video-sections of sociable interaction and music therapy were watched repeatedly to develop rich descriptions of the behavior and expressions constituting the themes. These descriptions were condensed into thematic items operationalizing verbal and nonverbal expressions (Bakeman & Quera, 2012; Chorney et al., 2015; Heath et al., 2010). In team meetings including the main supervisor (IHN), co-supervisor (MH), and me, these initial items were discussed extensively while watching video-examples to critically cross-check my interpretations and the content of the themes (Malterud, 2001). An expert in music therapy interventions was consulted for opinions about the initial items and their relevance and comprehensiveness for people living with dementia.

3.4.3 Field-testing of OWLS 1.0

A manual was developed to instruct coders about the definition of the codes with examples of behaviors included in the different items (Bakeman & Quera, 2012). Two research assistants were trained in the coding scheme using Noldus Observer XT ® (Noldus Information Technology, 2015). After the initial agreement reached >80% between coders, reliability-assessments of 20% of the video-material were conducted. During reliability-assessment and training, the research assistants gave feedback on the comprehensiveness and feasibility of the instrument and the interpretation of items (Terwee et al., 2018). The field-testing of the first version of OWLS was conducted in the sample described in Paper 2.

3.4.4 Statistical approach

The psychometric properties of the instrument were assessed using the COSMIN-definitions of the hypothesis testing of construct validity and responsiveness (Prinsen et al., 2018). Inter/intra-rater reliability was assessed with Cohen’ s Kappa, and measurement error was assessed as a percentage of agreement, in line with standards for nominal leveled instruments (Mokkink et al., 2020). Pearson’ s correlations were performed in R (R Core Team, 2020) and RStudio (RStudio Team, 2020), and inter/intra-rater reliability and percentage of agreement was calculated in Noldus Observer XT ® (Noldus Information Technology, 2015).

The following a priori hypotheses were tested according to:

1. Inter-rater reliability and measurement agreement
    - Kappa values >.80
    - % agreement >80%
2. Construct validity assessed through hypothesis testing
    - OWLS correlates higher with VNVIS-CR subscale Nonverbal Interaction than total VNVIS-CR score
    - OWLS correlates between .30-.70 with VNVIS-CR subscale Nonverbal Interaction
3. Responsiveness assessed through hypothesis testing
    - The LRR effect size (change score) of OWLS from the intervention will correlate ≥ .30 with self-reported change score of 'happiness' (VAMS)
    - The LRR effect size (change score) of OWLS from the intervention will correlate ≥ .30 with the change score of neuropsychiatric symptoms (NPI-Q)

3.4.5 Expert assessment of OWLS 1.0

Two focus groups involving experts were held to assess OWLS further, constituting a member-check of the content validity of the instrument (Brod et al., 2009; Collins, 2015; Malterud, 2001). The focus group interview was divided into two parts based on a semi-structured interview guide. The interview guide is available in Appendix B. First, open-ended questions were asked to investigate the opinions of the experts about what well-being in dementia is and how it could be observed. Next, the draft of the instrument was provided together with questions about the expert assessment of the content validity of the instrument according to COSMIN; the relevance, comprehensiveness, and comprehensibility of the content (Terwee et al., 2018). The experts were asked to evaluate how well the instrument would fit the psychosocial interventions they used in their daily clinical work with people living with dementia.

The recorded interviews were transcribed verbatim. The analysis was first conducted by me and the main supervisor (IHN) individually before we compared our records. We searched for the participants underlying understanding of what well-being is and how it is observed, as well as expressions related to the three topics of content validity (Braun & Clarke, 2006; Brod et al., 2009; Terwee et al., 2018). Keywords and examples were identified to further elaborate the operationalizations (Vogt et al., 2004).

3.4.6 Field-testing of OWLS 2.0

The suggested changes and refinement from the focus groups led to a revision that I field-tested in 10% of the video material from the intervention study to ensure the coding was feasible and valid. Intra-rater reliability for video coding in two separate weeks was assessed in Noldus Observer XT ® (Noldus Information Technology, 2015) against quality criteria of SCRIBE (Kappa values >.80 and agreement >80%). It was decided that the development process had reached a satisfactory point of saturation, and the concluding revision of OWLS was finalized (Chorney et al., 2015).

3.4.7 Reflexivity and validity of the mixed methods approach

The research criteria for the qualitative methods in the instrument development study were based on the Critical Appraisal Skills Programme (CASP) approach recommended by NICE (2015; Appendix H). Qualitative inquiries are always situated, and the validity- or audit-checks can increase the trustworthiness of the research process (Collins, 2015). The qualitative observation includes an interpretation of other people's expressions, and the engagement with the data started with my subjective clinical impression. In qualitative research, reflexivity is a central aim. It is defined as "*an attitude of attending systematically to the context of knowledge construction, especially to the effect of the researcher, at every step of the research process*" (Malterud, 2001; p 484). Several steps were taken to critically evaluate the subjectivity of the interpretations of the observations. The development phase utilized team-based discussions in the fieldwork-phase, assessing reliability between coders in the field-test, testing correlations with other instruments against a priori hypotheses, conducting focus groups to allow other persons to evaluate the instrument, and repeating field-testing and assessing coder agreement. The mixing of methods through building on the results from the former phase, intended to increase the validity of the content of OWLS. Shortcomings or problems detected in one phase would be addressed in the next phase. Additionally, the thematic analysis of the focus group interviews was conducted separately by me and the main supervisor (IHN) before they were discussed in the team.

3.4.8 Ethical considerations

The participants in the field-test were the same as in Paper 1, and the ethical considerations regarding this sample can be read in section 3.3.5. All participants in the focus groups gave written informed consent. Data handling was approved in advance by the Norwegian Centre for Research Data (ID: 489856). Ideally, people with dementia and their caregivers would also have been included in focus groups for elaborating on content validity (Terwee et al., 2018; Vogt et al., 2004), but because of the COVID-19 pandemic this inclusion was not an option. Existing qualitative research concerning the conceptualization of well-being from the perspectives of people living with dementia was consulted to compensate for this reality.

4. Summary of results

4.1 Summary of results Paper 1

Paper 1 is entitled *"Assessing momentary well-being in people living with dementia: A systematic review of observational instruments."* The systematic search detected 9474 records. After hand searches, 25 records were added. Full-text assessments were conducted for 87 articles including 47 instruments. A total of 36 records were eligible for inclusion, reporting on the development and psychometric properties of 22 instruments.

The included instruments were conceptualized as a) observations of emotions, b) observations of positive behavioral expressions, and c) observations of engagement. Table 5 includes the 22 instruments with their abbreviations and key references.

	Table 5 – instruments included in the review	
	Instrument	**Key references**
Emotions	**(Emotion) Facial Action Coding System (EMFACS)**	Ekman and Friesen (1978) Asplund et al. (1991) Asplund et al. (1995)
	The Maximally Discriminative Facial Movement Coding System (MAX)	Izard (1979) Izard (1995) Magai et al. (1996)
	Observed Emotion Rating Scale (OERS)	Lawton et al. (1996); Lawton et al. (1999)
	The Apparent Emotion Rating Scale (AER)	Snyder et al. (1998)
	Observable Displays of Affect (ODAS)	Beck et al. (2002); Vogelpohl and Beck (1997)
Positive expressions	**Dementia Care Mapping version 8 (DCM-8)**	Bradford Dementia Group (2005); Brooker and Surr (2006)
	Positive Response Schedule (PRS)	Perrin (1997)
	Activity in Context and Time (ACT)	Wood (2005)
	Greater Cincinnati Chapter Well-Being Observational Tool (GCWBT)	Kinney and Rentz (2005)

	Scripps Modified Greater Cincinnati Chapter Well-being Observation Tool (SM-GCWBT)	Lokon et al. (2019); Sauer et al. (2016)
	AwareCare	Clare et al. (2012)
	Behavior, Engagement and Affect Measure (BEAM)	Casey et al. (2014)
	Maastricht Electronic Daily Life Observational tool (MEDLO-tool)	de Boer et al. (2016)
	COMMUNI-CARE	Lopez et al. (2016)
	QUALIDEM for intensive longitudinal assessment (QUALIDEM-ILA)	Junge et al. (2020)
Engagement	**Menorah Park Engagement Scale (MPES)**	Judge et al. (2000)
	Observational Measurement of Engagement (OME)	Cohen-Mansfield et al. (2009)
	Music in Dementia Assessment Scales (MiDAS)	McDermott et al. (2014); McDermott et al. (2015)
	Video Coding – Incorporating Observed Emotion (VC-IOE)	Jones et al. (2015)
	Engagement of a Person with Dementia Scale (EPWDS)	Jones et al. (2018)
	Ethographic and Laban Inspired Coding System of Engagement/Evidence-Based Model of Engagement-Related-Behavior (ELICSE/EMODEB)	Perugia et al. (2018)
	Music therapy engagement scale (MTED)	Tan et al. (2019)

Content validity was the best-documented measurement property. Seventeen of 22 instruments were rated as “good”, and 11 of these were supported by high-quality evidence. Seventeen instruments were assessed for intra-rater reliability, but most had “conflicting” evidence. Sixteen instruments were assessed for construct validity, most with conflicting evidence or low-quality evidence. Evidence of structural validity, internal consistency, intra-rater reliability, test-retest reliability, measurement error, and measurement invariance were evaluated in less than half of the instruments, most with either low ratings or low evidence quality. No instruments assessed responsiveness or cross-cultural validity specific for the dementia population.

Although no specific instruments could be recommended, feasibility and interpretability were assessed for the instruments with best content validity. This assessment indicated the most promising results for PRS, AwareCare, BEAM, QUALIDEM-ILA, ELICSE, EPWDS, and MTED.

Paper 1 is under review.

4.2 Summary of results Paper 2

Paper 2 is entitled *"Observing music therapy in dementia: Repeated single-case studies assessing well-being and sociable interaction"*.

4.2.1 Assessment of reliability, validity, and treatment fidelity

Inter-rater agreement was good to excellent for VNVIS-CR ($\kappa = .92$, 92% agreement) and OWLS ($\kappa = 0.82$, 84% agreement). The variability of scores and mostly absent baseline and intervention trend supported the choice of LRR as a valid effect size parameter in this context. Assessment of treatment logs showed that the music-therapist led sessions were focused on musical elements. There was low adherence to complete the musical sessions led by the caregiver collateral.

4.2.2 Meta-analyses

The robust cluster meta-analysis of the single-case effect sizes showed a 48% increase in well-being (OWLS) from baseline to intervention ((*LRR* =0.39, 95% CI = [0.28 – 0.51], SE=0.05, t(9.88)=7.49, p=<0.001***, $I^2 = 86.03$, $\tau^2 = 0.04$). For sociable interaction (VNVIS-CR ratio) the increase from baseline to intervention was 32% (*LRR* = 0.28, 95% CI = [0.05 – 0.50], SE=0.10, t(9.97)=2.78, p=<0.02*, $I^2 = 90.62$, $\tau^2 = 0.16$). Heterogeneity of the results were high and was mostly not explainable. Meta-regression analyses did not support an increasing effect over time, and depression did not predict treatment effect. However, for the subscale nonverbal sociable interaction (VNVIS-CR nonverbal ratio) dementia severity predicted intervention effect. People with moderate dementia showed a 93% increase of nonverbal sociable interaction during music therapy (LRR = 0.65, $I^2 = 91.21$, $\tau^2 = 0.18$).

4.2.3 Analysis of behavioral content

For OWLS-items, the behavioral items that changed the most from baseline to intervention were “enjoyment” (45.9%), “happiness” (23.2%), “express identity” (36.5%), “relationship” (34.3%), and “positive feedback” (14.1%). For VNVIS-CR, the behavioral items that changed the most from baseline to intervention were “positive affect” (33%), “calm/relaxed” (20.9%), “appears aloof” (-19.4%), “vocalize negative affect” (-10.6%) and “responds to questions” (-30.9%). Assessing the nonverbal items of VNVIS-CR in the subgroup of moderate dementia showed even larger differences between the two phases and included changes >10% in the additional items; “looks at partner” (12.7%), “appears interested” (11%), “positive affect” (46.2%), “calm/relaxed” (30.7%), “appears aloof” (-27%), “stares into space” (-10%), and “vocalize negative affect” (-14.6%). For the remaining items of both scales the relative frequencies changed less than 10%.

4.2.4 Measures from pre- to post

All the pre-post session measures with VAMS were heavily skewed and could not be analyzed with hierarchical linear models as planned. Thus, the Wilcoxon signed-rank test of the pre-post session measures on group level was conducted. The analysis found an increase in self-reported positive emotions after the music therapy sessions (median pre= 46, median post= 63.5, V= 2640.5, p = <0.001***). The median difference was close to 20 points. Negative self-reported emotions were decreasing as well (median pre = 7, median post = 4, V = 38538, p = <0.001***) with a substantial floor effect.

The measures from pre- to post-intervention were not normal distributed. Wilcoxon signed-rank test showed no change on the long-term self-reported well-being (QoL-AD; median pre = 22, median post = 27, V = 15, p = 0.72) and no change in caregiver burden (RSS; median pre = 24, median post = 27, V = 21.5, p = 0.67). The neuropsychiatric symptoms (NPI-Q) decreased from before to after the intervention (median pre = 12, median post = 8, V = 36, p = 0.014*).

This paper is accepted in *Clinical Gerontologist.*

4.3 Summary of results Paper 3

Paper 3 is entitled *"Development of the Observable Well-being in Living with Dementia-Scale (OWLS): a measure for well-being during psychosocial and creative interventions".* The review of relevant well-being theory and instruments in Paper 1 and fieldwork using the video graphed music therapy sessions led to the preliminary version of OWLS. Items included in the field test of Paper 2 were "attention", "initiative/response," "happiness," "joking," "enjoyment," "mastery," "self-confidence," "reminiscence," "positive feedback," and "relationship".

Through the field-testing in Paper 2, preliminary results of psychometric properties were acquired. Inter-rater agreement between the trained coders and the main coder was good with a Cohen's Kappa of 0.82 ($N_o = 417$; CI = [0.72 – 0.89], p= <0.001***) and a percentage of agreement of 84% (N_o= 417; range 77 – 88%).

The construct validity was supported with significant Pearson's correlations of OWLS and VNVIS-CR total ratio (r (2197) = .37, p = <0.001***), and higher correlations with the VNVIS-CR nonverbal ratio (r(2197) = .65, p = <0.00***1). More severe dementia correlated positively with well-being (r(30) = .56, p = <0.001***), demonstrating that increasing dementia severity did not lower the potential of high well-being scores.

The responsiveness of OWLS was supported by significant Pearson's correlations with change-scores of NPI-Q (r (30) = .42, p = 0.017*) and the item "happy" from VAMS (r (30) = .33, p = 0.037*).

The focus group analysis provided support for the relevance of most items. "Self-confidence" was removed because it was evaluated as complicated to detect during observation. "Reminiscence" was rephrased as "express identity". Three items were suggested and field-tested ("calm/relaxed," "significant emotional experiences," and "participation").

The last field-test, including 10% of the video material, led to the revision of OWLS where eight items were included. These were “attention,” “initiative/response,” “calm/relaxed,” “happiness,” “enjoyment,” “express identity,” “mastery,” and “relationship”. Intra-rater agreement of OWLS was calculated for coding one week apart with excellent scores ($N_o = 220$, $\kappa = .98$, CI = [.94 – 1], $p = <0.0001$***; percentage of agreement 98%, range 94 – 100%).

The full OWLS-scale is available in Paper 3. This paper is submitted.

5. Discussion

The overarching objective of this thesis was to contribute to the knowledge base of well-being for people living with dementia, approached in three papers. Improving the prospect of living well with dementia is important. The biopsychosocial model of dementia outlines how psychosocial interventions can target tractable factors to optimize function and support well-being.

First, Paper 1 systematically reviewed and evaluated observational well-being instruments applicable for assessing psychosocial interventions for people living with dementia. This review contributed with structured evaluations of measurement properties, quality of evidence at instrument level, and finally, the clinical utility of the specific instruments. The review can guide researchers and practitioners in the process of choosing the relevant instrument best fit for the purpose of assessment.

Next, the potential of music therapy to target tractable psychosocial factors and facilitate positive experiences in dementia was investigated in a single-case study reported in Paper 2. The music therapy was tailored to the individual preferences and building on relevant autobiographic music. During music therapy, observed well-being and self-rated positive emotions were increased, targeting important individual consequences of dementia. Additionally, to target important relational consequences of dementia, the music therapy included family caregivers. Using single-case designs that provided the required assessment sensitivity in this context, these results support that music therapy is a valuable intervention approach for home-dwelling people living with dementia.

Lastly, the aim of developing an observational measure appropriate for assessing and comparing observable well-being during music therapy and regular social interaction was approached in Paper 3. The development procedures established acceptable content validity, promising preliminary psychometric properties, and adequate clinical utility of OWLS.

The three papers will be discussed in terms of how their findings contribute to and resonate with existing knowledge in the following sections, including implications for future research. The methodological strengths- and weaknesses, validity, trustworthiness, or legitimation of the results are discussed. These sections are followed by central ethical considerations, particularly regarding informed consent issues. Finally, the clinical implications of the research are summarized, with a conclusion of how this thesis contributes to the research field.

5.1 Observing momentary well-being in dementia

New research- and public health initiatives demand a specific focus on positive outcomes to optimize dementia interventions and care (Dowson et al., 2019; Reilly et al., 2020; WHO, 2017), and Paper 1 contributes by complementing previous systematizations of positive outcomes in dementia. The review offers a comprehensive overview and analysis of instrument-specific characteristics and addresses common methodological problems. Additionally, general- and instrument-specific recommendations are provided, contributing to the further development of the research field, and supporting future implementation of systematic observational well-being measures.

In Paper 1, important gaps in the knowledge base were identified. This included lack of data on responsiveness, measurement invariance, and cross-cultural validity across instruments. An important step for conducting valid quantitative intervention research is to choose measurement instruments with optimal validity for the specific population and context. The measurement properties of the instrument have implications for the validity of conclusions from intervention studies and systematic reviews of research as well (Mokkink et al., 2017). Several new instruments have been published following the most recent review of observational well-being measures for dementia (Algar et al., 2016). Additionally, further assessment of the recommended instrument (GCWBT) given by Algar et al. (2016) has demonstrated that the structural validity of this instrument was flawed (Gross et al., 2015; Lokon et al., 2019; Sauer et al., 2016).

The most important quality of any *observational* measure is to ensure that the inferences and interpretations regarding the observed persons are valid and trustworthy. For assessment of well-being, this implies that the interpretation reflects the inner state of the other person as accurately as possible (Algar et al., 2016; Bakeman & Quera, 2011). COSMIN recognizes content validity as the most important measurement property (Terwee et al., 2018). Eleven of the investigated instruments were rated as having acceptable content validity supported by high evidence quality. In most cases, the content validity was established in *one* study only, mostly supported by a strong theoretical foundation, assessment by professional experts, and adequate field-testing. However, inviting people with dementia or their family caregivers to elaborate on which elements are relevant and most comprehensively reflect well-being could strengthen an instruments' trustworthiness for inferring about inner states (Brod et al., 2009; Vogt et al., 2004). Such inquiries were only investigated for four instruments (AwareCare, Clare et al., 2012; EPDWS, Jones et al., 2018; QUALIDEM-ILA, Junge et al., 2020; MiDAS, McDermott et al., 2015).

Another way of investigating the ability of the instruments to validly infer about inner states would be through assessment of statistical correlations with self-reported instruments or other instruments known to validly detect well-being. The construct validity was acceptable for nine instruments, but there was a major lack of studies with trustworthy quality. Thus, future inquiries need to be carefully designed, including hypotheses postulated a priori with specified expected correlations (Prinsen et al., 2018). It is recommended to include more self-reported outcomes when applicable, and instruments measuring well-being on a momentary and not trait-like level. Especially since state and trait well-being is likely to differ (Cohen-Mansfield, 2011; Curyto et al., 2008).

Using ecological momentary assessment (EMA) has some clear advantages. It may detect mechanisms of change, which is an important issue that needs to be researched in psychosocial interventions for dementia (McDermott et al., 2019). Such mechanisms can be assessed by investigating changes and covariates as they unfold (Ayres & Gast, 2009; Bakeman & Quera, 2011). The high responsiveness and clinical sensitivity possible through ecological momentary designs are more likely to detect significant changes that is of importance for people living with dementia than the lower sensitivity of end-of-treatment assessment (Kishita et al., 2020; Shiffman et al., 2008). This being said, ecological momentary assessment approaches using *self-report* would ideally be the best way to learn about momentary variations in well-being, but this is not always an option because of the progression of cognitive impairment in dementia. Observation can also better detect a broader specter of clinically important data during psychosocial interventions that standard self-reporting instruments may miss or leave out (Algar et al., 2016). Using neutral observers seems like a feasible alternative, as proxy-ratings from family or professional caregivers are found to systematically report well-being lower than self-reports (Ferring & Boll, 2010; Schulz et al., 2013).

5.2 Positive changes during music therapy

In Paper 2, the primary hypotheses of a momentary increase in well-being (OWLS) and sociable interaction (VNVIS-CR) during the music therapy sessions were supported. The hypotheses of increased self-reported positive emotions (VAMS) after music therapy from pre- to post-sessions were also supported. The statistically significant decrease in self-reported negative emotions (VAMS) from pre- to post-sessions was influenced by a significant floor effect which is interpreted as reflecting a general infrequency of self-reported negative emotions. This result is in line with the common skewness of negative emotions found in Paper 1 as well. As opposed to the hypotheses, self-reported long-term well-being (QoL-AD) and caregiver-reported burden (RSS), was stable from pre- to post-intervention. However, the neuropsychiatric symptoms (NPI-Q) decreased significantly from pre- to post-session and showed a more favorable shift than our hypothesized anticipation of them remaining stable.

5.2.1 Individually tailored music therapy

The 48% increase in momentary well-being resonates well with current meta-analyses of music therapy, where a small increase in well-being or QoL is found at the end of treatment (van der Steen et al., 2018; Zhang et al., 2017). In line with the biopsychosocial model of dementia (Spector & Orrell, 2010), it seems that music therapy may accelerate some amenable factors that can facilitate optimal functioning for people with dementia. Some elements that could have accounted for this positive change is the tailoring of the intervention to the person's (and the dyad's) musical history and identity (Dowson et al., 2019; Lord et al., 2020), the strong emphasis on enjoyable activities (O'Rourke, Duggleby, et al., 2015) and the resource-oriented focus of the intervention that facilitates empowerment and independence of the participants (Bjørkløf et al., 2019; Holopainen et al., 2018). The amenable factors in music therapy outlined in the TMCM match well with the factors described in the biopsychosocial model and suggest music therapy reinforce well-being through fostering positive emotions, provide an including social environments, with optimal mental and sensory stimulation that support independence and function (Brancatisano et al., 2020; Spector & Orrell, 2010; Särkämö, 2018).

The specific behavioral expressions of well-being that changed from baseline to intervention at a group level seem to reflect clinically relevant outcomes and needs identified by people living with dementia. Enjoyment and happiness resonate well with meaningful and pleasurable activities demanded by people living with dementia (Miranda-Castillo et al., 2013; Reilly et al., 2020; Øksnebjerg et al., 2018). Expressed identity resonates well with "a sense of who you are" (Reilly et al., 2020). Relationships and positive feedback resonate well with facilitating communication and continued connection towards significant others (Reilly et al., 2020; Von Kutzleben et al., 2012).

5.2.2 Music therapy in a relational context

The 32% increase in sociable behavior towards significant others during music therapy supports that music therapy can be of value for both dyad members. In comparable intervention studies, the effect of music therapy on social behavior is mostly operationalized as a decrease in agitation or other types of disruptive behavior (Abraha et al., 2017; van der Steen et al., 2018). While the decrease in these symptoms is important, several triggers of negative emotions and experiences are associated with intrinsic stimuli that may be less modifiable by psychosocial interventions (Kales et al., 2015; Kolanowski et al., 2017). Examples of these can be pain, infections, or perceptual disturbances. The behavioral expressions that changed most from baseline to intervention were the sociable, positive expressions, which highlights the importance of assessing sociable and not only non-sociable behavior (Dowson et al., 2019; van der Steen et al., 2018). The two non-sociable behaviors that decreased the most were "appears aloof" and vocalizing negative affect. Adding the increase in positive affect and calm/relaxed expressions, suggests music therapy facilitated emotional availability in the participants. The behavioral changes that may follow dementia, where apathy is one of the most common symptoms (Steinberg et al., 2008), influence the capacity to interact with the environment. Thus, the 93% increase of the nonverbal sociable interaction during music therapy for people with moderately severe dementia may have important clinical implications for their interpersonal interactions with their surroundings. In a biopsychosocial frame, music therapy seems to expand functioning in social relations and facilitate closeness (Spector & Orrell, 2010). Reasonable mechanisms that can make social capacities available are that music captures attention and prevents apathy. The music allows for nonverbal and meaningful communication not dependent on spoken language (Brancatisano et al., 2020; Dowlen et al., 2018; Wadham et al., 2016).

While we did not find a decrease in caregiver burden at the end of treatment, the decrease in the level of neuropsychiatric symptoms suggests some general effect following the 10-week intervention. Some important clinical implications can be drawn from this. First, the level of BPSD is related to caregiver well-being and burden, as is a lack of mutuality between caregiver and care-receiver (Cheng, 2017;

Livingston et al., 2017). Next, caregiver burden predicts institutionalization of the care-receiver (Cheng, 2017). Enabling people living with dementia to remain in their homes for as long as possible has benefits for the individual level of well-being (Holopainen et al., 2018), as well as positive economic effects (Livingston et al., 2017). Thus, further investigating if music therapy can positively affect care giver burden through lowering symptoms of BPBS is of great clinical importance. Especially as most people with dementia live at home, and informal caregivers spend numerous hours caring for their close ones (Ydstebø et al., 2020).

5.2.3 Heterogeneity

The results from the meta-analyses were highly heterogeneous, and meta-regression only identified one significant predictor in the sub-analysis for nonverbal sociable interaction. Thus, the improvement from baseline to intervention varied largely, within sessions case-wise and between cases. One possible explanation of the variability is the fluctuating presence of BPSD in dementia (Kales et al., 2015), as well as daily variations in emotional well-being (Kolanowski et al., 2007). The "depression"-predictor in the meta-regression was not momentary but reflected the initial depression score from NPI-Q. A better approach to analyze predictors could be to assess symptoms present on the same day. Assessing individual co-variates within the person living with dementia, co-occurring elements of the music therapy, or co-occurring interactional elements from significant others, is valuable in future studies to identify mechanisms and predictors of treatment effect.

5.3 The measurement properties and feasibility of OWLS

The purpose of developing OWLS was to approach the lack of content validity in existing observational instruments for assessing interventions with a strong nonverbal component such as music therapy. The existing instruments relied too much on verbal expressions. Our overarching aim in developing the coding-scheme was to establish an acceptable content validity for our target group and context. All the OWLS items have nonverbal *and* verbal indicators, and only one indicator is required for the item to be scored as "present."

5.3.1 Measurement properties

Relevant measurement properties are available in the COSMIN taxonomy from Mokkink et al. (2010), as presented in Paper 3, to evaluate the qualities of OWLS. The COSMIN quality appraisal and risk of bias assessment conducted in Paper 1 is also applicable and useful to compare OWLS to the existing instruments examined in Paper 1 (Mokkink et al., 2020; Prinsen et al., 2018; Terwee et al., 2018).

Regarding content validity, the systematic COSMIN-assessment would lead to rating “good” content validity with high evidence quality according to the criteria, including evidence of comprehensiveness and relevance established through literature review, fieldwork and expert focus groups, and comprehensibility investigated through team discussions and assessment in the expert focus group (Terwee et al, 2018).

Assessing structural validity and internal consistency would not apply, as OWLS is a formative instrument. Reliability and measurement error would be rated as “good,” according to quality criteria (>.70). However, it would be assessed as low-quality evidence in GRADE because of imprecision (low sample size, Mokkink et al., 2020). Further assessment of OWLS could elaborate reliability assessment through item-specific with Kappa, or reliability of the total score through ICC (Mokkink et al., 2017).

Criterion validity is not applicable, as there is no gold-standard instrument for comparison (Mokkink et al., 2017). For hypothesis testing of construct validity and responsiveness, results met “good” quality criteria. Results were following >75% of the hypotheses, but evidence quality is downgraded to low because of imprecision here as well (Prinsen et al., 2018). Compared to the existing observational instruments of well-being reviewed in Paper 1, OWLS is the only instrument formally assessing responsiveness. Another strength is the guide to interpreting clinically relevant changes when the effect size is calculated with LRR (Pustejovsky & Swan, 2018).

Regarding measurement invariance and cross-cultural validity, these are qualities that need to be further elaborated for OWLS. This issue holds for the other instruments assessed in Paper 1 as well. However, one pertinent asset of OWLS is the use of introverted and extroverted signs of enjoyment. A common cross-cultural variance in emotional expression is high versus low arousal (Lim, 2016). In OWLS, enjoyment can be expressed through closed eyes. a relaxed body, *or* energetic and engaged gestures.

5.3.2 Feasibility

The feasibility of OWLS can be investigated by the defining dimension of observational instruments outlined in the introduction (Cfr. Table 2), particularly the granularity, concreteness, and primary recording unit. The granularity of OWLS is somewhat fine-grained, as behaviors are coded in 30 seconds intervals. This approach enables small changes to be detected. The high granularity also increases the labor intensiveness of the instrument and suggests OWLS is best suited for research and video-recordings. This feedback also came from the expert group working in dementia care.

The concreteness of OWLS is on the lower side of the continuum, with socially based codes. Concrete codes are advocated as more objective (Ayres & Gast, 2009; Perugia et al., 2018), but the fieldwork of OWLS led us to conclude that social context was required and appeared to increase content validity. This conclusion may seem counterintuitive but can be elaborated through an example. ACT is an instrument on the concrete side of the continuum, and uses gaze to detect engagement (Wood, 2005). This code would have low relevance in our study. For example, some participants demonstrated engagement through the vigor, dynamics, and attunement of their voices while singing a song while their eyes were *closed*. Other highly concrete codes decrease content validity in people with dementia, as reported by Asplund et al. (1995). The highly concrete facial expression of emotion characterizing EMFACS did a poorer job of detecting emotions in people living with severe dementia than coding based on a holistic evaluation of an emotional expression in the persons face (Asplund et al., 1995). Facial and bodily expressions may be changed during the progression of dementia, implying that we need to think differently about the concreteness of the chosen indicators. Thus, if the codes may be reliably interpreted across observers, including social interpretation seems to increase the content validity in this context. Ultimately, using socially based codes seemed to increase the content validity and feasibility of OWLS.

OWLS was developed to provide a comprehensive estimation of the most important *observable* aspects of well-being in dementia. Nevertheless, well-being also consists of several *non-observable* aspects. Clarke et al. (2020) reported that other important elements might be self-confidence and life having meaning. These aspects seem better detected through self-report, and future studies utilizing OWLS are encouraged to triangulate the observations with validated self-report scales or qualitative interviews when applicable. This approach would also assess and maybe further legitimate using OWLS to draw inferences about inner states.

OWLS rely on partial interval sampling (Pustejovsky, 2019). An advantage of this primary recording unit is that any presence of an indicator is coded as an event, which may easier achieve good inter-rater reliability and measurement agreement as opposed to continuous recordings. However, partial interval sampling provides a less precise estimate of duration and intensity than other primary recording levels (Ayres & Gast, 2009). Instruments using continuous time-sampling, such as BEAM (Casey et al., 2014), can better estimate the frequency by reporting on the duration of the behavior of interest. Adding an evaluation of the intensity level of well-being expression could better distinguish between intervention effects. Nevertheless, adding complexity comes at a cost that may decrease reliability, increase time spent, and lower feasibility in many contexts (Bakeman & Quera, 2011).

OWLS does not assess negative expressions. This emphasis was a feasibility choice to avoid the skewness on item level that was problematic for several of the existing instruments reviewed in Paper 1. It is still important to monitor adverse effects when investigating health interventions (Tate et al., 2016). The use of logs in Paper 2 is an example of such monitoring and is recommended in future studies utilizing OWLS. Cross validating OWLS with other socially relevant instruments or approaches is of great importance. This is vital for any observational instrument or single-case assessment (Ledford & Gast, 2009; Tate et al., 2016).

5.4 Strengths and limitations

The central limitations of this thesis are discussed in the following section, for each paper individually. For Paper 1, this regards our adaptation of the RoB-checklist and the fairness of the ratings. For Paper 2, this regards sample size, internal and external validity, and explanatory power. For Paper 3, this regards the legitimation and inference quality of the mixed methods approach.

5.4.1 The methodological validity of Paper 1

Consensus-based standards for assessing the study-specific risk of bias and evaluating instrument-specific measurement properties and the trustworthiness of the results were utilized to increase the validity of the systematic review (Mokkink et al., 2020; Mokkink et al., 2017; Prinsen et al., 2018; Terwee et al., 2018). Furthermore, the data extracted were predetermined and pre-registered in PROSPERO, and the review was reported in line with the PRISMA Statement (Shamseer et al., 2015). Pre-defined inclusion criteria guided the screening of records from the search. The initial screening was completed by me alone and gave a risk of overlooking important records. However, the abstract screening was blinded between me and the main supervisor (IHN) and reached an acceptable 82% agreement.

The systematic assessment of content validity required some adaptation on the RoB-items from Terwee et al. (2018), as the validity of observational measures most of all relies on rigorous fieldwork and piloting the instrument in the field (Bakeman & Quera, 2011, 2012; Chorney et al., 2015). Content validity of observational measures cannot be investigated through focus groups or qualitative interviews alone. However, this is the consensus-based gold standard for establishing content validity of patient-rated instruments (Brod et al., 2009; Terwee et al., 2018). The modification of the RoB-items conducted by our research team were based on principles for developing observational coding schemes to minimize bias (Bakeman & Quera, 2011, 2012; Chorney et al., 2015). Adaptations of the COSMIN approach to other contexts are recommended by the COSMIN team when required (Mokkink et al., 2018). However, the adaptation has not been subject to the Delphi-procedures the original COSMIN-criteria are based on, where numerous experts give advice and feedback in several rounds.

No systematic approach is without limitations, and during the application of the adapted COSMIN-ratings, three concerns were central. The first concern regards how the COSMIN RoB-checklist quickly leads to low ratings, as the worst score counts in any domain assessed (Mokkink et al., 2017). If one instrument has three of five RoBs in one domain, and the other has only one RoB in the same domain, both may be rated as “inadequate” and downgraded to a low evidence quality with GRADE. All risks are equally weighed and may not fairly compare or distinguish between studies with one compared to several present RoBs.

The next concern regards the use of checklists for evaluating qualitative work when assessing content validity. For qualitative inquiries, evaluations traditionally rely on more holistic assessments of the research process. The COSMIN checklist does not fully grasp a thorough agenda, such as for example EPICURE (Stige et al., 2009). Thus, the COSMIN criteria will downgrade a study not recording and transcribing the qualitative interviews verbatim but do not address other important domains for the validity of qualitative research. Important subjects such as reflexivity, transparency or interpretive validity are omitted (Finlay & Gough, 2003; Stige et al., 2009; Whittemore et al., 2001). This background is the reason Paper 1 is described as a *QUANT*itative dominated inquiry in the introduction. The qualitative component is clearly present, as COSMIN encourages establishing content validity through qualitative methods. However, the COSMIN-approach seems to be developed within a quantitative tradition, with objectivity as the main criterion for generating and evaluating new knowledge. This attempt to mix qualitative and quantitative methods in a unified checklist is an important step to secure measurement instruments are informed by the “best available evidence,” but the attempt to evaluate the qualitative evidence within a quantitative tradition is a shortcoming for COSMIN to rigorously evaluate qualitative research.

The last concern regards the criteria of construct validity and responsiveness tested through hypothesis testing. Instruments compared to, and correlating with, only one other instrument will get a “good”-rating as the results would be following >75% of the hypotheses. Instruments investigating correlations with more than one instrument would originally get an “inadequate” rating if six out of nine correlations support the instrument (Prinsen et al., 2018). In the latter case, it seems evident that the construct validity is more thoroughly elaborated. Thus, the COSMIN quality criteria were adapted in Paper 1 implying that instruments in the latter category were rated as “conflicting” instead of “inadequate”. AwareCare (Clare et al., 2012) is a relevant example, where five of eight hypotheses were supported, but correlations with proxy rated QoL from family caregivers and staff were <.3. A plausible alternative explanation for this may be that the validity of the proxy-ratings is influenced by attributes of the proxy, as is found in several other studies (Ferring & Boll, 2010; Schulz et al., 2013). Thus, rating instruments as “conflicting” instead of “inadequate” provided a fairer comparison to instruments less elaborated.

Still, COSMIN is to date the most elaborated systematic approach for assessing health-related measurement instruments, and the approach is developing to cover other categories of health-related measurement approaches (Mokkink et al., 2020).

5.4.2 The methodological validity of Paper 2

Changes in pre-registered protocol

Paper 2 was pre-registered as a clinical trial, which is conducted to increase the validity of intervention studies. The outcome measures registered in the first round were later evaluated as not content valid for the context of music therapy when I piloted the coding-schemes in Noldus Observer XT (Noldus Information Technology, 2015). The choice to exchange the instrument CODEM (Kuemmel et al., 2014) with VNVIS-CR (Williams et al., 2017) allowed assessing nonverbal communication behavior in isolation. It removed the problem of relying on verbal expressions that were predominant in CODEM. Instead of using OERS (M. P Lawton et al., 1999), we developed and used OWLS mainly because OERS is predominated by items with negative emotional expressions. Significant floor-effects have hampered previous studies using OERS in the data analysis (Algar et al., 2016). While changing outcome measures can be interpreted as a source of bias, coding the observations with measures with low content validity for the specific context was a bigger threat to the trustworthiness of the results.

Utilizing the clinical sensitivity of single-case designs

Further research is vital to pinpoint how and when music therapy has a beneficial effect on people living with dementia. When standard RCT-designs from various reasons may be unfeasible, the time series methodology utilized in single-case designs enables testing changes of interventions implemented in a clearly defined time point, as well as comparing baseline with intervention phases (Bernal et al., 2016; Manolov & Moeyaert, 2017; Manolov & Vannest, 2019). As opposed to between-group designs, these within-case approaches are important to acquire better insight into how interventions work at an individual level, as group-level methods cannot conclude about intra-individual variability (Piccirillo et al., 2019). Additionally, the problem of detecting clinically relevant changes in dementia research encourages the use of fine-grained and sensitive instruments and designs (Schall et al., 2015). Both means are achievable in single-case designs and were regarded as methodological benefits in Paper 2.

Optimizing valid detection of clinically significant changes

While the single-case design showed favorable change during music therapy, the assessment of the validity of the results is always a central concern. An apparent strength of the design in Paper 2 is an optimal ecological validity. Acting as one’s own control is a meaningful comparison, omitting the problem of using between-group data to infer about individual subjects (Piccirillo et al., 2019). However, the explanatory power and generalizability of small group data are of concern when single-case designs are conducted, and reviews of music therapy and other psychosocial interventions in dementia have been concerned about the small sample sizes (Abraha et al., 2017; McDermott et al., 2019; van der Steen et al., 2018).

Two widely cited single-case design guidelines were used to assess the explanatory power and validity of the single-case design in Paper 2 (Kratochwill et al., 2012; Tate et al., 2016). Steps that were taken to increase internal validity was to use assessors independent of therapists, systematically manipulate the independent variable three times for each participant, and choose a statistical approach validly assessing the relevant parameters of our data (Kratochwill et al., 2012; Manolov et al., 2014; Tate et al., 2016). Assessing baseline and intervention trends suggested this was not a major issue in our data, supporting the internal validity. To further optimize the internal validity, attempts to blind or mask the assessors, conduct a randomization procedure such as in multiple-baseline designs, or other attempts to control for maturation or history are relevant opportunities (Manolov et al., 2014; Tate et al., 2016). This possibility could better have ruled out alternative explanations to the increase from baseline to intervention.

The steps that were taken to increase external validity were replicating the design in 11 cases and training and assessing the inter-rater reliability of three different coders to make sure sufficient agreement was established (Tate et al., 2016). Generalizability would increase if the study were replicated with similar findings on other research sites by other researchers (Kratochwill et al., 2012). The legitimacy of our observational inferences as reflecting the participants inner states could have been assessed by cross-validating the outcomes with self-report measures or qualitative interviews. The use of VAMS was planned to provide data fitted for a hierarchical linear model that would shed light on the levels of self-reported emotions throughout the 10-week intervention (Woltman et al., 2012). This approach would have assessed music therapy within cases and over time from the perspective of the participant. Unfortunately, the skewness of these data only allowed for a non-parametric analysis at the group level.

All the pre-post measures (VAMS, NPI-Q, QoL-AD, and RSS) were assessed within a quasi-experimental design, and it is important to be cautious about any inferences drawn. The explanatory power of these uncontrolled variables was low but triangulating the data can either support or weaken the plausibility of the results of the single-case design (Kratochwill et al., 2012). In this case, two of the four instruments suggested positive effects, and none suggested adverse effects.

5.4.3 The methodological validity of Paper 3

While all three papers are positioned within a mixed-methods approach, Paper 1 and 2 are parallel when using qualitative and quantitative approaches. Paper 3 is the only paper mixing qualitative and quantitative methods through the building approach (Cfr. Figure 3). The different qualitative and quantitative phases of developing the instrument in Paper 3 were first evaluated with corresponding separate quality criteria.

Evaluation against quality criteria of the fieldwork and focus groups

Evaluating the qualitative work against predefined criteria (NICE, 2015; Terwee et al., 2018) identifies some strengths and limitations.

As the fieldwork analysis was initially conducted by me independently, this required specific attention to the subjectivity and interpretations during the analysis was handled critically. Inviting the team to investigate the themes and central video segments representing them contributed to an important cross-check of the analysis (Malterud, 2001). Staying close to the observed phenomenon and use descriptive examples before condensing these descriptions to themes was an explicit intention to increase the credibility of the items (Smith et al., 2009). In hindsight, a more systematic approach to report on reflexive self-critique would have been advantageous for increasing the transparency of this analysis and the continuing reflexive process that was undertaken (Finlay & Gough, 2003). Sharing qualitative video transcripts in Paper 3 allowed the reader to evaluate the trustworthiness of the themes (Stige et al., 2009). Sharing the interview guide also increase transparency on the research method (NICE, 2015)

The initial analysis of the focus group was conducted by me and the main supervisor (IHN) independently before discussing the results in the team. This point is a strength advocated in the COSMIN-guidelines (Terwee et al., 2018). The focus group interviews added valuable information from the intended users of OWLS, which was a clear advantage to increase the feasibility of the instrument. However, further valuable information could have been acquired through focus groups where the experts coded observations with the instrument and gave feedback during coding. A central weakness was the lack of qualitative interviews including people living with dementia and their informal caregivers, which could have adjusted the OWLS items further and increased the trustworthiness of the content (Terwee et al., 2018).

Evaluation against quality criteria for the quantitative assessment

The evidence quality of the quantitative methods in the development of OWLS is discussed in section 5.1.3. The main threat to validity is the small sample size.

Legitimacy and inference quality of the exploratory sequential methodology

Mixing methods can add value to research when the strengths of one approach account for the limitations of the other (Creswell, 2015). The purpose of the methodological *building* was to address shortcomings in any of the phases through developing the instrument further (Fetters et al., 2013). The choice of methods was based on pragmatism, asking which method best could answer the objectives and research question at hand (Dures et al., 2010). Three specific threats to validity were identified in the methods chapter, which will be discussed here.

The first threat was quantitizing the qualitative themes from the fieldwork of Paper 3 (and 2). Qualitative descriptions are richer and have the advantage of anchoring the operationalizations close to the phenomenon in description before the standardization of the theme into a dichotomous code. Quantifying qualitative themes will always risk simplifying the richness of the qualitative inquiries (Fetters et al., 2013). However, using dichotomous codes was advantageous for comparing behavioral content in a single-case design that was highly sensitive to small changes, thus resolving a central methodological sensitivity issue for assessing psychosocial interventions and music therapy.

The second threat is mixing inductive and deductive logics that were built upon in the individual development phases. The theoretical review and fieldwork led to 10 themes with operationalizations for the initial development of items. Then, testing these items in the field-test (Paper 3) or intervention study (Paper 2) violates the deductive logic in quantitative traditions and quality standards of postulating hypotheses a priori (Collins, 2015). In mixed methods, using deductive and inductive logic in the same study is common, and using inductive explorative and deductive explanatory approaches is expected to be a cyclical process (Teddlie, 2010). Discovery in the inductive logic can be followed by justification from the deductive logic or vice versa, if the sampling procedure is valid (Morse, 2010; Teddlie, 2010). This discussion brings us to the third validity threat, which was the sampling procedures of the quantitative and qualitative assessment.

Carefully conducted sampling is important to avoid validity threats in mixed methods (Collins, 2015). Using an identical sample in the qualitative and quantitative part of a mixed-methods research is invalid for two reasons. The sample will be too large and not reflecting the phenomena of interest for qualitative studies, and the sample will be too small and biased for quantitative studies (Morse, 2010). In Paper 3, the qualitative and quantitative data were drawn from the same participants. In Paper 3, a purposive *sub-sample* was used in the qualitative fieldwork, where video segments were watched and significant sections reflecting well-being themes were identified and described (Braun & Clarke, 2006; Heath et al., 2010). Thus, it did not include the complete dataset, but only different segments from some of the participants until the same themes re-emerged. Reaching “saturation” is a quality criterion for qualitative inquiries (NICE), 2015; Terwee et al., 2018) and guided the sampling of the fieldwork. For the quantitative part of Paper 3, including *all* the observational data from the sample was required, as the presence and absence of the items were of interest. Thus, the sampling procedure had a clear purpose in the different parts of the development process. Hopefully the transparency allows the reader to conclude about the legitimacy of the sampling (Fetters et al., 2013).

5.5 Ethical issues

The clinical study and data handling followed formal external and internal registration procedures and evaluation (REK, NSD, RETTE), which included external evaluation or control of ethical issues and data handling outside the research group. The Norwegian Centre for Research Data approved the data handling plan for the focus group in Paper 3. Papers 2 and 3 were also registered in the System for Risk and compliance (RETTE) at the University of Bergen. All research data, including videos, were stored safely in the University of Bergen's secure system SAFE. The development of the research project was planned and conducted without directions from financial providers. Informed consent was collected in the projects where human participants were included (Paper 2 and 3). For the people living with dementia, this will be elaborated on in the next section.

The Health Research Act that regulates the inclusion of humans in research requires consent to be *"informed, voluntarily, explicitly and documented"* (Helseforskningsloven, 2009, §4-13). The Helsinki declaration advice the inclusion of vulnerable groups in research and demands that the risk and disadvantage of participations largely outweigh the advantages and gain the individual or other persons from the population of interest (World Medical Association, 2013).

The capacity to provide informed consent include four components; 1) Understand and remember information regarding diagnosis and treatment, 2) relate this information to one's personal condition, 3) reason, compare and choose between different treatment alternatives, and 4) communicate this choice of treatment (Moye et al., 2007). Reaching these stringent criteria can be challenging when including people with dementia in research, as the progression of dementia will impair several cognitive domains relevant for this capacity in moderate and more severe stages of dementia. While the legally informed consent may be impaired in some people in the more severe stages of dementia, the ethical recommendations advice always seeking the participants' *endorsement* in the research (World Medical Association, 2013). Moreover, several approaches and recommendations were assessed that describe how one can facilitate *the ability* for people with dementia to provide informed consent (Dewing, 2007; McKillop & Wilkinson, 2004; Moye et al., 2007; Warner et al., 2008). It is also a central ethical concern to not exclude people living with dementia from the important opportunities to express their opinions or contribute to research (Reid et al., 2018). In Paper 2 we included several steps to facilitate informed consent for the participants living with dementia. First, the informed consent form was written in plain language – the time from information to consent was short to facilitate the first, second and third element in the descriptions of informed consent by Moye et al. (2007).

Additionally, written information was provided. Time was given to think about the decision to participate, reminders of the intervention being a research project, explaining why the videos were taken, and viewing informed consent as a process instead of a one-time event (Dewing, 2007) to facilitate communication of one's own choice. Treating informed consent as a process also included the researchers and clinicians protecting the rights of the participants during the research (Moye et al., 2007).

If signs of withdrawal were brought to our attention by the music therapist or caregivers, this was discussed. Thus, the participants informed consent or endorsement was always sought. The caregiver participants were also included in the evaluation of the care-receivers endorsement and consent. The family caregiver was always present during the first meeting where informed consent was sought. They were encouraged to support the person with dementia in their decision making and discuss participation with them.

There is tension between protecting vulnerable groups in research and experiencing paternalism that decreases personal autonomy (Reid et al., 2018). Many individuals living with dementia have expressed an explicit demand to be included in research and states this is empowering and meaningful (McKillop & Wilkinson, 2004; Øksnebjerg et al., 2018). People in the early stages of dementia address that they do not want to be denied the right to participate in research despite some risks and that this is a choice the person wants to make themselves (Øksnebjerg et al., 2018). Moye et al. (2007) advocate that personal values should be considered when judging informed consent to individualize the assessment and facilitate decisions in line with the person's preferences.

All the participants with dementia in Paper 2 gave written consent to participate. Several safeguards were included to protect the well-being and rights of the participants. First, developing a contingency plan if adverse experiences or symptom development happened during the music therapy intervention. This plan was given to the participants and included written information and contact details about the psychologist or psychiatrist responsible for the outpatient treatment. Next, music therapists in Norway do not have formal authorization as healthcare professionals, and supervision from a clinical psychologist was available if concerns about the psychological health of the participants were present. Lastly, the assessment of symptoms before inclusion was conducted by a clinical psychologist, and exclusion criteria were severe psychotic symptoms and the presence of suicide ideation.

After collecting the research data in Paper 2, the major impression was that the participants were capable of, and interested in, participating in the research. The processual approach of informed consent assessment was helpful for balancing the protection and autonomy of the participants. Typical situations triggering skepticism and reminders of the volunteer nature of participation in the research project were in the video filmed music therapy sessions The participants mostly accepted video graphing after a conversation where concerns could be freely expressed and clarified.

The music therapists also raised a concern regarding the use of the VAMS instrument assessing self-reported emotions. In VAMS, participants are asked to rate the presence of six negative emotions and two positive emotions (Stern et al., 1997). The main emphasis on negative emotions was of concern and seemed to be experienced as intrusive by some participants. Thus, VAMS was omitted when this was the case. This balancing of validity of the research data and ethical concerns must necessarily favor the participants' protection. Newly developed instruments emphasizing more positive experiences would probably better have served the function of momentary self-reports, such as the Canterbury Wellbeing Scales (Strohmaier et al., 2021).

Lastly, one ethical concern raised after the Paper 2 and 3 was the validity of our observational research representing the "inner world" of the people living with dementia. In hindsight, supplementing the observations by conducting qualitative interviews with the participants in Paper 2 would have been preferable. Including people with dementia and family caregivers in induvial or focus group interviews assessing content validity for OWLS (Paper 3) would have been sought. However, this desire was not an option because of the COVID-19 pandemic.

5.6 Clinical implications and future directions

Following the implications and research recommendations already discussed in the sections above, some central clinical implications follow. The strong focus on decreasing BPSD and other challenges that has dominated research on psychosocial interventions for dementia is important, but not sufficient, to improve well-being (Clarke et al., 2020; Harrison et al., 2016; Stoner et al., 2019). Unlike psychotropic medication, many promising psychosocial interventions are unavailable for people living with dementia. We know that most psychotropic medication has limited effects and damaging side effects, but still they are often offered as the first choice (Livingston et al., 2017; Winblad et al., 2016). Several psychosocial interventions lack support when assessed in systematic reviews (Abraha et al., 2017; Livingston et al., 2017; McDermott et al., 2019). However, promising psychosocial interventions need to be assessed with sensitive measurement designs and sensitive measures with sufficient measurement properties. The lack of knowledge about measurement invariance in existing instruments has implications for the conclusions of clinical inquiries. Particularly, dementia severity and dependency can be confounding variables masking the treatment effect of interventions. Lack of measurement sensitivity risks depriving people living with dementia access to important interventions that could be helpful to live well with dementia. Ecologic momentary assessment is one such sensitive approach, and single-case research is a sensitive design option.

This thesis identified music therapy as one approach targeting central tractable factors described in the biopsychosocial model of dementia (Spector & Orrell, 2010). However, the large variability in effect within and between cases, implies that we need to know more about covariates and moderators to better tailor the interventions to the individual's needs. Under which circumstances does music therapy target tractable factors, and when is it not? Future research needs to address these questions.

5.7 Conclusion

Dementia has individual, relational and societal consequences and is a highly prevalent condition in the aging population worldwide. We need to know more about tractable factors enabling people and caregivers to live well with dementia. One approach is to use observational measures to assess the effect of a variety of promising psychosocial interventions. This thesis contributed with a structured review of the available observational well-being measures. Many instruments had promising content validity but had conflicting evidence of construct validity and reliability and lacked cross-cultural validity, responsiveness, and measurement invariance. Music therapy is one promising approach for alleviating the individual and relational consequences of dementia. The single-case designs utilized in this thesis demonstrated clinically important momentary changes in well-being and nonverbal sociable interactions during music therapy for dyads. Additionally, the development of an observational instrument contributed to an approach that can compare music therapy fairly to social interaction, identifying potential advantages in either activity. OWLS needs to be further assessed in larger samples, other contexts, and validated independently by other researchers. Future research should focus on specifying under which circumstances music therapy can be advantageous, particularly in larger samples with home-dwelling people living with dementia.

6. References

Ablitt, A., Jones, G. V., & Muers, J. (2009). Living with dementia: a systematic review of the influence of relationship factors. *Ageing and mental health, 13*, 497-511. https://doi.org/10.1080/13607860902774436

Abraha, I., Rimland, J. M., Trotta, F. M., Dell'Aquila, G., Cruz-Jentoft, A., Petrovic, M., Gudmundsson, A., Soiza, R., O'Mahony, D., Guaita, A., & Cherubini, A. (2017). Systematic review of systematic reviews of non-pharmacological interventions to treat behavioural disturbances in older patients with dementia. The SENATOR-OnTop series. *BMJ Open, 7*(3), e012759-e012759. https://doi.org/10.1136/bmjopen-2016-012759

Algar, K., Woods, R. T., & Windle, G. (2016). Measuring the quality of life and well-being of people with dementia: A review of observational measures [Neurological Disorders & Brain Damage 3297]. *Dementia: The International Journal of Social Research and Practice, 15*(4), 832-857. https://doi.org/http://dx.doi.org/10.1177/1471301214540163

Alzheimer's Disease International, & Patterson, C. (2018). World Alzheimer's Report 2018 - The state of the art of dementia research: New frontiers. 1-48. https://www.alzint.org/resource/world-alzheimer-report-2018/

Alzheimer's Society. (2021). The dementia guide: Living well after your diagnosis. www.alzheimers.org.uk/dementiaguide

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders V. https://doi.org/10.1176/appi.books.9780890425596

Asplund, K., Jansson, L., & Norberg, A. (1995). Facial expressions of patients with dementia: A comparison of two methods of interpretation [Neurological Disorders & Brain Damage 3297]. *International Psychogeriatrics, 7*(4), 527-534. https://doi.org/http://dx.doi.org/10.1017/S1041610295002262

Asplund, K., Norberg, A., Adolfsson, R., & Waxman, H. M. (1991). Facial expressions in severely demented patients: A stimulus-response study of four patients with dementia of the Alzheimer type [Neurological Disorders & Brain Damage 3297]. *International Journal of Geriatric Psychiatry, 6*(8), 599-606. https://doi.org/http://dx.doi.org/10.1002/gps.930060809

Ayres, K., & Gast, D. L. (2009). Dependent Measures and Measurement Procedures. In J. R. Ledford & D. L. Gast (Eds.), *Single Subject Research Methodology in Behavioral Sciences* (pp. 129-165). London: Taylor & Francis Group. https://doi.org/10.4324/9780203877937-7

Baird, A., & Thompson, W. F. (2018). The Impact of Music on the Self in Dementia. *J Alzheimers Dis, 61*(3), 827-841. https://doi.org/10.3233/JAD-170737

Bakeman, R., & Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge University Press. https://doi.org/10.1017/CBO9781139017343

Bakeman, R., & Quera, V. (2012). Behavioral observation. In H. Cooper, M. Paul, L. L. Debra, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research medhods in psychology* (pp. 207-225). American Psychological Association. https://doi.org/10.1037/13619-013

Baker, F. A., Grocke, D., & Pachana, N. A. (2012). Connecting through music: A study of a spousal caregiver-directed music intervention designed to prolong fulfilling relationships in couples where one person has dementia. *Australian journal of music therapy, 23*(2012), 4-21.

Bartlett, R., Windemuth-Wolfson, L., Oliver, K., & Dening, T. (2017). Suffering with dementia: the other side of "living well". *International Psychogeriatrics, 29*(2), 177-179. https://doi.org/10.1017/S104161021600199X

Beck, C. K., Vogelpohl, T. S., Rasin, J. H., Uriri, J. T., O'Sullivan, P., Walls, R., Phillips, R., & Baldwin, B. (2002). Effects of behavioral interventions on disruptive behavior and affect in demented nursing home residents. *Nursing Research, 51*(4), 219-228. https://doi.org/10.1097/00006199-200207000-00002

Beerens, H. C., de Boer, B., Zwakhalen, S. M. G., Tan, F. E. S., Ruwaard, D., Hamers, J. P. H., & Verbeek, H. (2016). The association between aspects of daily life and quality of life of people with dementia living in long-term care facilities: a momentary assessment study. *International Psychogeriatrics, 28*(8), 1323-1331. https://doi.org/10.1017/S1041610216000466

Bernal, J. L., Cummins, S., & Gasparrini, A. (2016). Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyw098

Bjørkløf, G. H., Helvik, A.-S., Ibsen, T. L., Telenius, E. W., Grov, E. K., & Eriksen, S. (2019). Balancing the struggle to live with dementia: a systematic meta-synthesis of coping. *BMC Geriatrics, 19*(1), 295-295. https://doi.org/10.1186/s12877-019-1306-9

Bowling, A., Rowe, G., Adams, S., Sands, P., Samsi, K., Crane, M., Joly, L., & Manthorpe, J. (2015). Quality of life in dementia: A systematically conducted narrative review of dementia-specific measurement scales [Neurological Disorders & Brain Damage 3297]. *Aging & Mental Health, 19*(1), 13-31. https://doi.org/http://dx.doi.org/10.1080/13607863.2014.915923

Bradford Dementia Group. (2005). *DCM 8 User's Manual: The DCM Method*. University of Bradford.

Brancatisano, O., Baird, A., & Thompson, W. F. (2020). Why is music therapeutic for neurological disorders? The Therapeutic Music Capacities Model. *Neuroscience and Biobehavioral Reviews, 112*, 600-615. https://doi.org/10.1016/j.neubiorev.2020.02.008

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

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*(1), 25-35. https://doi.org/10.1093/geront/39.1.25

Brod, M., Tesler, L. E., & Christensen, T. L. (2009). Qualitative research and content validity: developing best practices based on science and experience. *Quality of Life Research, 18*(9), 1263-1278. https://doi.org/10.1007/s11136-009-9540-9

Brooker, D., & Surr, C. (2006). Dementia Care Mapping (DCM): Initial validation of DCM 8 in UK field trials. *International Journal of Geriatric Psychiatry, 21*(11), 1018-1025. https://doi.org/10.1002/gps.1600

Burns, R. A. (2017). Psychosocial Well-Being. In N. A. Pachana (Ed.), *Encyclopedia of Geropsychology* (pp. 1977-1984). Springer. https://doi.org/10.1007/978-981-287-082-7_251

Casey, A.-N., Low, L.-F., Goodenough, B., Fletcher, J., & Brodaty, H. (2014, 2014-07). Computer-assisted direct observation of behavioral agitation, engagement, and affect in long-term care residents. *Journal of the American Medical Directors Association, 15*(7), 514-520. https://doi.org/10.1016/j.jamda.2014.03.006

Cheng, S.-T. (2017). Dementia caregiver burden: a research update and critical analysis. *Current Psychiatry Reports, 19*(9), 1-8. https://doi.org/10.1007/s11920-017-0818-2

Cheng, S.-T., Lam, L., & Kwok, T. (2013). Neuropsychiatric symptom clusters of alzheimer disease in Hong Kong Chinese: Correlates with caregiver burden and depression. *American Journal of Geriatric Psychiatry, 21*(10), 1029-1037. https://doi.org/10.1016/j.jagp.2013.01.041

Chorney, J. M., McMurtry, C. M., Chambers, C. T., & Bakeman, R. (2015). Developing and modifying behavioral coding schemes in pediatric psychology: A practical guide. *Journal of Pediatric Psychology, 40*(1), 154-164. https://doi.org/10.1093/jpepsy/jsu099

Clair, A. A. (2002). The effects of music therapy on engagement in family caregiver and care receiver couples with dementia. *American Journal of Alzheimers Disease and Other Dementias, 17*(5), 286-290. https://doi.org/10.1177/153331750201700505

Clare, L., Whitaker, R., Quinn, C., Jelley, H., Hoare, Z., Woods, B., Downs, M., & Wilson, B. (2012). AwareCare: Development and validation of an observational measure of awareness in people with severe dementia [Neuropsychological Assessment 2225]. *Neuropsychological Rehabilitation, 22*(1), 113-133. https://doi.org/http://dx.doi.org/10.1080/09602011.2011.640467

Clark, I. N., Tamplin, J. D., & Baker, F. A. (2018). Community-dwelling people living with dementia and their family caregivers experience enhanced relationships and feelings of well-being following therapeutic group singing: A qualitative thematic analysis. *Frontiers in Psychology, 9*, 1332-1332. https://doi.org/10.3389/fpsyg.2018.01332

Clarke, C., Woods, B., Moniz-Cook, E., Mountain, G., Øksnebjerg, L., Chattat, R., Diaz, A., Gove, D., Vernooij-Dassen, M., & Wolverson, E. (2020). Measuring the well-being of people with dementia: a conceptual scoping review. *Health and Quality of Life Outcomes, 18*(1), 1-249. https://doi.org/10.1186/s12955-020-01440-x

Clements-Cortes, A., & Bartel, L. (2018). Are we doing more than we know? Possible mechanisms of Response to music therapy. *Frontiers in Medicine, 5*, 255-255. https://doi.org/10.3389/fmed.2018.00255

Cohen-Mansfield, J. (2011). The Shifting Baseline Theory of Well-Being: Lessons from across the Aging Spectrum. In L. W. Poon & J. Cohen-Mansfield (Eds.), *Understanding Well-Being in the Oldest Old* (pp. 46-64). Cambridge University Press.

Cohen-Mansfield, J., Dakheel-Ali, M., & Marx, M. S. (2009). Engagement in persons with dementia: The concept and its measurement [Neurological Disorders & Brain Damage 3297]. *The American Journal of Geriatric Psychiatry, 17*(4), 299-307. https://doi.org/http://dx.doi.org/10.1097/JGP.0b013e31818f3a52

Collins, K. M. T. (2015). Validity in Multimethod and Mixed Research. In S. N. Hesse-Biber & R. B. Johnson (Eds.), *The oxford handbook of multimethod and mixed methods research inquiry* (pp. 240-256). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199933624.013.17

Cooper, C., Mukadam, N., Katona, C., Lyketsos, C., Blazer, D., Ames, D., Rabins, P., Brodaty, H., de Mendonça Lima, C., & Livingston, G. (2013). Systematic Review of the effectiveness of pharmacologic interventions to improve quality of life and well-being in people with dementia. *American Journal of Geriatric Psychiatry, 21*(2), 173-183. https://doi.org/10.1016/j.jagp.2012.10.018

Creswell, J. W. (2015). Revisiting Mixed Methods and Advancing Scientific Practices. In S. N. Hesse-Biber & R. B. Johnson (Eds.), *The oxford handbook of multimethod and mixed methods research inquiry* (pp. 57-71). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199933624.013.39

Cuddy, L. L., Sikka, R., Silveira, K., Bai, S., & Vanstone, A. (2017). Music-evoked autobiographical memories (MEAMs) in Alzheimer disease: Evidence for a positivity effect. *Cogent psychology, 4*(1). https://doi.org/10.1080/23311908.2016.1277578

Cuddy, L. L., Sikka, R., & Vanstone, A. (2015). Preservation of musical memory and engagement in healthy aging and Alzheimer's disease. *Annals of the New York Academy if Science, 1337*(1), 223-231. https://doi.org/10.1111/nyas.12617

Curyto, K., Haitsma, K., & Vriesman, D. (2008). Direct observation of behavior: A review of current measures for use with older adults with dementia. *Research in Gerontological Nursing, 1*(1), 52-76. https://doi.org/10.3928/19404921-20080101-02

de Boer, B., Beerens, H. C., Zwakhalen, S. M. G., Tan, F. E. S., Hamers, J. P. H., & Verbeek, H. (2016). Daily lives of residents with dementia in nursing homes: development of the Maastricht electronic daily life observation tool. *International Psychogeriatrics, 28*(8), 1333-1343. https://doi.org/10.1017/S1041610216000478

de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in Medicine: A Practical Guide*. Cambridge University Press. https://doi.org/10.1017/CBO9780511996214

Dewing, J. (2007). Participatory research: A method for process content with persons who have dementia. *Dementia: The International Journal of Social Research and Practice, 6*, 11-25. https://doi.org/10.1177/1471301207075625

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 103-157.

Diener, E. (2006). Guidelines for National Indicators of Subjective Well-Being and Ill-Being. *Applied Research in Quality of Life, 1*(2), 151-157. https://doi.org/10.1007/s11482-006-9007-x

Dodge, R., Daly, A. P., Huyton, J., & Sanders, L. D. (2012). The challenge of defining wellbeing. *International journal of wellbeing, 2*(3), 222-235. https://doi.org/10.5502/ijw.v2.i3.4

Dowlen, R., Keady, J., Milligan, C., Swarbrick, C., Ponsillo, N., Geddes, L., & Riley, B. (2018). The personal benefits of musicking for people living with dementia: a thematic synthesis of the qualitative literature. *Arts & Health, 10*(3), 197-212. https://doi.org/10.1080/17533015.2017.1370718

Dowson, B., McDermott, O., & Schneider, J. (2019). What indicators have been used to evaluate the impact of music on the health and wellbeing of people with dementia? A review using meta-narrative methods. *Maturitas, 127*, 26-34. https://doi.org/http://dx.doi.org/10.1016/j.maturitas.2019.06.001

Dröes, R. M., Chattat, R., Diaz, A., Gove, D., Graff, M., Murphy, K., Verbeek, H., Vernooij-Dassen, M., Clare, L., Johannessen, A., Roes, M., Verhey, F., Charras, K., & The Interdem Social Health Taskforce. (2016). Social health and dementia: a European consensus on the operationalization of the concept and directions for research and practice. https://doi.org/10.1080/13607863.2016.1254596

Dures, E., Rumsey, N., Morris, M., & Gleeson, K. (2010, Mar). Mixed methods in health psychology: Theoretical and practical considerations of the third paradigm. *Journal of Health Psychology, 16*(2), 332-341. https://doi.org/10.1177/1359105310377537

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System (FACS): Manual.* Consulting Psychologist Press.

El Haj, M., Clément, S., Fasotti, L., & Allain, P. (2013). Effects of music on autobiographical verbal narration in Alzheimer's disease. *Journal of neurolinguistics, 26*(6), 691-700. https://doi.org/10.1016/j.jneuroling.2013.06.001

El Haj, M., Postal, V., & Allain, P. (2012). Music enhances autobiographical memory in mild alzheimer's disease. *Educational Gerontology, 38*(1), 30-41. https://doi.org/10.1080/03601277.2010.515897

Engstrom, G., Hammar, L., Williams, C., & Gotell, E. (2011). The impact of singing in caring for a person with dementia: Single case analysis of video recorded sessions. *Music and Medicine, 3*(2), 95-101. https://doi.org/10.1177/1943862110386048

Feast, A., Moniz-Cook, E., Stoner, C., Charlesworth, G., & Orrell, M. (2016). A systematic review of the relationship between behavioral and psychological symptoms (BPSD) and caregiver well-being. *International Psychogeriatrics, 28*(11), 1761-1774. https://doi.org/10.1017/S1041610216000922

Ferring, D., & Boll, T. (2010). Subjective well-being in older adults: Current state and gasp of research. In L. Bovenberg & A. Van Soest (Eds.), *Aging, health and pensions in Europe. An economic and social policy perspective* (pp. 173-205). Palgrave Macmillan.

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Service Research, 48*(6pt2), 2134-2156. https://doi.org/10.1111/1475-6773.12117

Finlay, L., & Gough, B. (2003). *Reflexivity: A practical guide for researchers in health and social sciences*. John Wiley & Sons Inc. http://ebookcentral.proquest.com/lib/bergen-ebooks/detail.action?docID=351401

Fisher, Z., & Tipton, E. (2017). *robumeta: Robust Variance Meta-Regression.* In

Fusar-Poli, L., Bieleninik, Ł., Brondino, N., Chen, X.-J., & Gold, C. (2018). The effect of music therapy on cognitive functions in patients with dementia: A systematic review and meta-analysis. *Aging and Mental Health, 22*(9), 1103-1112. https://doi.org/10.1080/13607863.2017.1348474

Gaugler, J. E., Bain, L. J., Mitchell, L., Finlay, J., Fazio, S., Jutkowitz, E., Banerjee, S., Butrum, K., Fazio, S., Gaugler, J., Gitlin, L., Hodgson, N., Kallmyer, B., Le Meyer, O., Logsdon, R., Maslow, K., & Zimmerman, S. (2019). Reconsidering frameworks of Alzheimer's dementia when assessing psychosocial outcomes. *Alzheimer's & Dementia: Translational Research & Clinical Interventions, 5*(1), 388-397. https://doi.org/10.1016/j.trci.2019.02.008

Gjøra, L., Strand, B. H., Bergh, S., Borza, T., Brækhus, A., Engedal, K. A., Johannessen, A., Kvello-Alme, M., Krokstad, S., Livingston, G., Matthews, F. E., Myrstad, C., Skjellegrind, H., Thingstad, P., Aakhus, E., Aam, S., & Selbæk, G. (2021). Current and future prevalence estimates of mild cognitive impairment, dementia, and its subtypes in a population-based sample of people 70 years and Ooder in Norway: The HUNT Study. *Journal of Alzheimer's Disease*. https://doi.org/10.3233/JAD-201275

Gold, C., Solli, H. P., Krüger, V., & Lie, S. A. (2009). Dose–response relationship in music therapy for people with serious mental disorders: Systematic review and meta-analysis. *Clinical Psychology Review, 29*, 193-207. https://doi.org/doi:10.1016/j.cpr.2009.01.001

Graham, J. E., & Bassett, R. (2006). Reciprocal relations: The recognition and co-construction of caring with Alzheimer's disease. *Journal of Aging Studies, 20*(4), 335-349. https://doi.org/10.1016/j.jaging.2005.12.003

Greene, J. G., Gardiner, M., & Timbury, G. C. (1982). Measuring behavioral disturbance of elderly demented patiens in the community and its effects on relatives: a factor analytic study. *Age and ageing, 11*, 121-126.

Gross, S. M., Danilova, D., Vandehey, M. A., & Diekhoff, G. M. (2015). Creativity and dementia: Does artistic activity affect well-being beyond the art class? [Health Psychology Testing 2226]. *Dementia: The International Journal of Social Research and Practice, 14*(1), 27-46. https://doi.org/http://dx.doi.org/10.1177/1471301213488899

Harding, A. J. E., Morbey, H., Ahmed, F., Opdebeeck, C., Lasrado, R., Williamson, P. R., Swarbrick, C., Leroi, I., Challis, D., Hellstrom, I., Burns, A., Keady, J., & Reilly, S. T. (2019, 2019/03/27). What is important to people living with dementia?: the 'long-list' of outcome items in the development of a core outcome set for use in the evaluation of non-pharmacological community-based health and social care interventions. *BMC Geriatrics, 19*(1), 94. https://doi.org/10.1186/s12877-019-1103-5

Harrison, J. K., Noel-Storr, A. H., Demeyere, N., Reynish, E. L., & Quinn, T. J. (2016). Outcomes measures in a decade of dementia and mild cognitive impairment trials. *Alzheimer's Research and Therapy, 8*(1), 48-48. https://doi.org/10.1186/s13195-016-0216-8

Heath, C., Hindmarsh, J., & Luff, P. (2010). Analysing video: developing preliminary observations. In *Video in Qualitative Research: Analysing Social Interaction in Everyday Life* (pp. 61-85). SAGE Publications, Inc. https://doi.org/https://www.doi.org/10.4135/9781526435385

Hedges, L. V. (2019). Stochastically dependent effect sizes In C. Harris, L. V. Hedges, & J. C. Valentine (Eds.), *Handbook of Research Synthesis and Meta-Analysis* (3 ed., pp. 281-298). New York: Russel Sage Foundation. https://doi.org/10.7758/9781610441384.23

Helsedirektoratet. (2019). *Demensplan 2025*. Helsedirektoratet. https://www.helsedirektoratet.no/tema/demens/demensplan-2020

Holopainen, A., Siltanen, H., Pohjanvuori, A., Mäkisalo-Ropponen, M., & Okkonen, E. (2018). Factors sssociated with the quality of life of people with dementia and with quality of life-improving interventions: Scoping review. *Dementia: The International Journal of Social Research and Practice, 18*(4), 1507-1537. https://doi.org/10.1177/1471301217716725

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry, 240*, 565 - 572. https://doi.org/10.1192/bjp.140.6.566

Hughes, L. J., Farina, N., Page, T. E., Tabet, N., & Banerjee, S. (2019). Psychometric properties and feasibility of use of dementia specific quality of life instruments for use in care settings: a systematic review. *International Psychogeriatrics*, 1-15. https://doi.org/10.1017/S1041610218002259

Istvandity, L. (2017). Combining music and reminiscence therapy interventions for wellbeing in elderly populations: A systematic review. *Complementary Therapies in Clinical Practice, 28*, 18-25. https://doi.org/10.1016/j.ctcp.2017.03.003

Izard, C. E. (1979). *The maximally discriminative facial movement coding system (MAX)*. University of Delaware.

Izard, C. E. (1995). *The maximally discriminative facial movement coding system (MAX) (Rev. ed)*. University of Delaware.

Jing, W., Willis, R., & Feng, Z. (2016). Factors influencing quality of life of elderly people with dementia and care implications: A systematic review. *Archives of Gerontolgy and Geriatics, 66*, 23-41. https://doi.org/10.1016/j.archger.2016.04.009

Johnson, B. R., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher, 33*(7), 14-26. https://doi.org/10.3102/0013189X033007014

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of mixed methods research, 1*(2), 112-133. https://doi.org/10.1177/1558689806298224

Jones, C., Sung, B., & Moyle, W. (2015). Assessing engagement in people with dementia: A new approach to assessment using video analysis. *Archives of Psychiatric Nursing, 29*(6), 377-382. https://doi.org/10.1016/j.apnu.2015.06.019

Jones, C., Sung, B., & Moyle, W. (2018). Engagement of a Person With Dementia Scale: Establishing content validity and psychometric properties [Clinical Psychological Testing 2224]. *Journal of Advanced Nursing, 74*(9), 2227-2240. https://doi.org/http://dx.doi.org/10.1111/jan.13717

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*(1), 42-46. https://doi.org/10.1177/153331750001500105

Junge, S., Gellert, P., O'Sullivan, J. L., Moller, S., Voigt-Antons, J.-N., Kuhlmey, A., & Nordheim, J. (2020). Quality of life in people with dementia living in nursing homes: Validation of an eight-item version of the qualidem for intensive longitudinal assessment [Psychological & Physical Disorders 3200]. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 29*, 1721–1730. https://doi.org/http://dx.doi.org/10.1007/s11136-020-02418-4

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Phys Life Rev, 10*(3), 235-266. https://doi.org/10.1016/j.plrev.2013.05.008

Kales, H. C., Gitlin, L. N., & Lyketsos, C. G. (2015). Assessment and management of behavioral and psychological symptoms of dementia. *BMJ: British Medical Journal, 350*, 1-16. https://doi.org/10.1136/bmj.h369

Kaufer, D. I., Cummings, J. L., Ketchel, P., Smith, V., MacMillan, A., Shelley, T., Lopez, O. L., & DeKosky, S. T. (2000). Validation of the NPI-Q, a brief clinical form of the neuropsychiatric inventory. *The Journal of Neuropsychiatry and Clinical Neurosciences, 12*, 233-239. https://doi.org/10.1176/jnp.12.2.233

Kinney, J. M., & Rentz, C. A. (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*(4), 220-227. https://doi.org/10.1177/153331750502000406

Kishita, N., Backhouse, T., & Mioshi, E. (2020). Nonpharmacological interventions to improve depression, anxiety, and Quality of Life (QoL) in people with sementia: An overview of systematic reviews. *Journal of Geriatric Psychiatry and Neurology, 33*(1), 28-41. https://doi.org/10.1177/0891988719856690

Kitwood, T. (1997). *Dementia reconsidered: The person comes first*. Buckingham: Open University Press.

Kolanowski, A., Boltz, M., Galik, E., Gitlin, L. N., Kales, H. C., Resnick, B., Van Haitsma, K. S., Knehans, A., Sutterlin, J. E., Sefcik, J. S., Liu, W., Petrovsky, D. V., Massimo, L., Gilmore-Bykovskyi, A., MacAndrew, M., Brewster, G., Nalls, V., Jao, Y.-L., Duffort, N., & Scerpella, D. (2017). Determinants of behavioral and psychological symptoms of dementia: A scoping review of the evidence. *Nursing Outlook, 65*(5), 515-529. https://doi.org/10.1016/j.outlook.2017.06.006

Kolanowski, A., Fortinsky, R. H., Calkins, M., Devanand, D. P., Gould, E., Heller, T., Hodgson, N. A., Kales, H. C., Kaye, J., Lyketsos, C., Resnick, B., Schicker, M., & Zimmerman, S. (2018). Advancing research on care needs and supportive approaches for persons with dementia: Recommendations and rationale. *Journal of Americal Medical Directors Association (JAMDA), 19*(12), 1047-1053. https://doi.org/10.1016/j.jamda.2018.07.005

Kolanowski, A., Hoffman, L., & Hofer, S. M. (2007). Concordance of self-report and informant assessment of emotional well-being in nursing home residents with dementia. *Journals of Gerontology Series B: Psychological Sciences & Social Sciences, 62*(1), 20-27. https://doi.org/10.1093/geronb/62.1.P20

Kolanowski, A. M. (1999). An overview of the need-driven dementia-compromised behavior model. *Journal of Gerontological Nursing, 25*(9), 7-9. https://doi.org/10.3928/0098-9134-19990901-05

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2012). Single-Case Intervention Research Design Standards. *Remedial and special education, 34*(1), 26-38. https://doi.org/10.1177/0741932512452794

Kuemmel, A., Haberstroh, J., & Pantel, J. (2014). CODEM instrument: Developing a tool to assess communication behavior in dementia. *The Journal of Gerontopsychology and Geriatric Psychiatry, 27*(1), 23-31. https://doi.org/10.1024/1662-9647/a000100

Kwasky, A., Harrison, B., & Whall, A. (2010). Quality of Life and Dementia. *Alzheimer's Care Today, 11*(3), 186-195. https://doi.org/10.1097/ACQ.0b013e3181ec00e1

Lawton, M. P. (1994). Quality of life in Alzheimer's disease. *Alzheimer Disease & Associated Disorders, 8*, 138-150. https://doi.org/10.1097/00002093-199404000-00015

Lawton, M. P. (1997). Assessing quality of life in Alzheimer disease research. *Alzheimer Disease and Associated Disorders, 11*(SUPPL. 6), 91-99. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed6&NEWS=N&AN=28029827

Lawton, M. P., Van Haitsma, K., & Klapper, J. (1996, 1996). Observed affect in nursing home residents with Alzheimer's disease. *The journals of gerontology. Series B, Psychological sciences and social sciences, 51*(1), P3-P14. https://doi.org/10.1093/geronb/51b.1.p3

Lawton, M. P., Van Haitsma, K., & Klapper, J. A. (1999). *Observed Emotion Rating Scale*. Retrieved 21. december from www.abrahamsoncenter.org/PRI

Lawton, M. P., Van Haitsma, 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*(1), 69-81.

Ledford, J. R., & Gast, D. L. (Eds.). (2009). *Single subject research methodology in behavioral sciences*. Taylor & Francis Group. https://ebookcentral-proquest-com.pva.uib.no/lib/bergen-ebooks/detail.action?docID=455501.

Leung, D. K. Y., Chan, W. C., Spector, A., & Wong, G. H. Y. (2021). Prevalence of depression, anxiety, and apathy symptoms across dementia stages: A systematic review and meta-analysis. *International Journal of Geriatric Psychiatry*. https://doi.org/10.1002/gps.5556

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ, 339*, b2700. https://doi.org/10.1136/bmj.b2700

Lim, N. (2016, 2016/06/01/). Cultural differences in emotion: differences in emotional arousal level between the East and the West. *Integrative Medicine*

Research, 5(2), 105-109. https://doi.org/https://doi.org/10.1016/j.imr.2016.03.004

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., Ballard, C., Banerjee, S., Burns, A., Cohen-Mansfield, J., Cooper, C., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Larson, E. B., Ritchie, K., Rockwood, K., Sampson, E. L., Samus, Q., Schneider, L. S., Selbæk, G., Teri, L., & Mukadam, N. (2017). Dementia prevention, intervention, and care. *Lancet, 390*, 2673-2734. https://doi.org/10.1016/S0140-6736(17)31363-6

Lobbia, A., Carbone, E., Faggian, S., Gardini, S., Piras, F., Spector, A., & Borella, E. (2019). The efficacy of Cognitive Stimulation Therapy (CST) for people with mild-to-moderate dementia: A review. *European psychologist, 24*(3), 257-277. https://doi.org/10.1027/1016-9040/a000342

Logsdon, R. G., Gibbons, L. E., M., M. S., & Teri, L. (1999). Quality of Life in Alzheimer's disease. *Journal of Mental Health and Aging, 5*, 21-32.

Logsdon, R. G., Gibbons, L. E., McCurry, S. M., & Teri, L. (2002). Assessing quality of life in older adults with cognitive impairment. *Psychosomatic Medicine, 64*(3), 510-519. https://doi.org/10.1097/00006842-200205000-00016

Lokon, E., Sauer, P. E., & Li, Y. (2019). Activities in dementia care: A comparative assessment of activity types [Health & Mental Health Services 3370]. *Dementia: The International Journal of Social Research and Practice, 18*(2), 471-489. https://doi.org/http://dx.doi.org/10.1177/1471301216680890

Lopez, J. J. B., Bolivar, J. C. C., & Perez, M. S. (2016). COMMUNI-CARE: Assessment tool for reactions and behaviours of patients with dementia in a multisensory stimulation environment [Clinical Psychological Testing 2224]. *Dementia: The International Journal of Social Research and Practice, 15*(4), 526-538. https://doi.org/http://dx.doi.org/10.1177/1471301214528346

Lord, K., Beresford-Dent, J., Rapaport, P., Burton, A., Leverton, M., Walters, K., Lang, I., Downs, M., Manthorpe, J., Boex, S., Jackson, J., Ogden, M., & Cooper, C. (2020). Developing the New Interventions for independence in Dementia Study (NIDUS) theoretical model for supporting people to live well with dementia at home for longer: a systematic review of theoretical models and Randomised Controlled Trial evidence. *Social Psychiatry and Psychiatric Epidemiology, 55*(1), 1-14. https://doi.org/10.1007/s00127-019-01784-w

Lyketsos, C. G., Carrillo, M. C., Ryan, J. M., Khachaturian, A. S., Trzepacz, P., Amatniek, J., Cedarbaum, J., Brashear, R., & Miller, D. S. (2011). Neuropsychiatric symptoms in Alzheimer’s disease. *Alzheimers Dement., 7*(5), 532-539. https://doi.org/10.1016/j.jalz.2011.05.2410

Ma, H.-H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points Exceeding the Median. *Behavior Modifcation, 30*(5), 598-617. https://doi.org/10.1177/0145445504272974

Madsø, K. G., & Nordhus, I. H. (2021). Implementation of Quality of Life Assessment in Long Term Care. *International Psychogeriatrics*, 1-8. https://doi.org/10.1017/S1041610221000983

Magai, C., Cohen, C., Gomberg, D., Malatesta, C., & Culver, C. (1996, 1996-09). Emotional expression during mid- to late-stage dementia. *International Psychogeriatrics, 8*(3), 383-395. http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=105822295&site=ehost-live

Malterud, K. (2001). Qualitative research: standards, challenges, and guidelines. *The Lancet, 358*, 483-488. https://doi.org/10.1016/S0140-6736(01)05627-6.

Manolov, R., Gast, D. L., Perdices, M., & Evans, J. J. (2014). Single-case experimental designs: Reflections on conduct and analysis. *Neuropsychological Rehabilitation: Single-case Experimental Design Methodology, 24*(3-4), 634-660. https://doi.org/10.1080/09602011.2014.903199

Manolov, R., & Moeyaert, M. (2017). How can single-case data be analyzed? Software resources, tutorial, and reflections on analysis. *Behavior Modifcation, 41*(2), 179-228. https://doi.org/10.1177/0145445516664307

Manolov, R., & Solanas, A. (2013). A comparison of mean phase difference and generalized least squares for analyzing single-case data. *Journal of School Psychology, 51*(2), 201-215. https://doi.org/10.1016/j.jsp.2012.12.005

Manolov, R., & Vannest, K. J. (2019). A Visual Aid and Objective Rule Encompassing the Data Features of Visual Analysis. *Behavior Modification*, 145445519854323-145445519854323. https://doi.org/10.1177/0145445519854323

Mark, M. M. (2015). Mixed and Multimethods in Predominantly Quantitative Studies, Especially Experiments and Quasi-Experiments. In C. Teddlie & A. Tashakkori (Eds.), *Mixed methods in social and behavioral research* (pp. 21-41). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199933624.013.15

Martyr, A., Nelis, S. M., Quinn, C., Wu, Y.-T., Lamont, R. A., Henderson, C., Clarke, R., Hindle, J. V., Thom, J. M., Jones, I. R., Morris, R. G., Rusted, J. M., Victor, C. R., & Clare, L. (2018). Living well with dementia: a systematic review and correlational meta-analysis of factors associated with quality of life, well-being and life satisfaction in people with dementia. *Psychological Medicine, 48*(13), 2130-2139. https://doi.org/10.1017/S0033291718000405

McDermott, O., Charlesworth, G., Hogervorst, E., Stoner, C., Moniz-Cook, E., Spector, A., Csipke, E., & Orrell, M. (2019). Psychosocial interventions for people with dementia: a synthesis of systematic reviews. *Aging and Mental Health, 23*(4), 393-403. https://doi.org/10.1080/13607863.2017.1423031

McDermott, O., Crellin, N., Ridder, H. M., & Orrell, M. (2013). Music therapy in dementia: A narative synthesis systematic review. *International Journal of Geriatric Psychiatry, 28*, 781-794. https://doi.org/10.1002/gps.3895

McDermott, O., Orgeta, V., Ridder, H. M., & Orrell, M. (2014). A preliminary psychometric evaluation of Music in Dementia Assessment Scales (MiDAS). *International Psychogeriatrics, 26*(6), 1011-1019. https://doi.org/10.1017/S1041610214000180

McDermott, O., Orrell, M., & Ridder, H. M. (2015). The development of Music in Dementia Assessment Scales (MiDAS). *Nordic Journal of Music Therapy, 24*(3), 232-251. https://doi.org/10.1080/08098131.2014.907333

McKeown, J., Clarke, A., Ingleton, C., & Repper, J. (2010). Actively involving people with dementia in qualitative research. *Journal of Clinical Nursing, 19*, 1935-1943. https://doi.org/10.1111/j.1365-2702.2009.03136.x

McKillop, J., & Wilkinson, H. (2004, Jun). Make it easy on yourself! Advice to researchers from someone with dementia on being interviewed. *Dementia: The International Journal of Social Research and Practice, 3*, 117-125. https://doi.org/10.1177/1471301204042332

Miranda-Castillo, C., Woods, B., & Orrell, M. (2013). The needs of people with dementia living at home from user, caregiver and professional perspectives: a cross-sectional survey. *BMC Health Service Research, 13*(1), 43-43. https://doi.org/10.1186/1472-6963-13-43

Missotten, P., Dupuis, G., & Adam, S. (2016). Dementia-specific quality of life instruments: a conceptual analysis. *International Psychogeriatrics, 28*(8), 1245. https://doi.org/10.1017/S1041610216000417

Mittelman, M. S., & Papayannopoulou, P. M. (2018). The Unforgettables: a chorus for people with dementia with their family members and friends. *International Psychogeriatrics, 30*(6), 779-789. https://doi.org/10.1017/S1041610217001867

Mokkink, L., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L., de Vet, H. C., & Terwee, C. (2018). *COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual*. COSMIN. https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf

Mokkink, L. B., Boers, M., van der Vleuten, C. P. M., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2020). COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Medical Research Methodology, 20*(1), 293-293. https://doi.org/10.1186/s12874-020-01179-5

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2017). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research, 27*(5), 1171-1179. https://doi.org/10.1007/s11136-017-1765-4

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology, 63*(7), 737-745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology, 43*, 2412-2414.

Morse, J. (2010). Procedures and Practice of Mixed Method Design: Maintaining Control, Rigor, and Complexity. In A. Tashakkori & C. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (2 ed., pp. 339-352). SAGE Publications, Inc. https://doi.org/10.4135/9781506335193.n14

Moye, J., Karel, M. J., Edelstein, B., Hicken, B., Armesto, J. C., & Gurrera, R. J. (2007). Assessment of Capacity to Consent to Treatment: Challenges, the "ACCT" approach, future directions. *Clinical Gerontologist, 31*(3), 37-66. https://doi.org/10.1080/07317110802072140

National Institute for Health and Care Excellence (NICE). (2015). Appendix H: Appraisal checklists, evidence tables, GRADE and economic profiles. In NICE (Ed.), *Clinical Guideline Development Methods: The Guidelines Manual 2015*. https://www.nice.org.uk/process/pmg20/resources/appendix-h-appraisal-checklists-evidence-tables-grade-and-economic-profiles-pdf-8779777885

Noldus Information Technology. (2015). *Noldus Observer XT, 12.5.* In Noldus Information Technology. www.noldus.com.

Norton, M. C., Piercy, K. W., Rabins, P. V., Green, R. C., Breitner, J. C. S., Østbye, T., Corcoran, C., Welsh-Bohmer, K. A., Lyketsos, C. G., & Tschanz, J. T. (2009). Caregiver–recipient closeness and symptom progression in Alzheimer disease. The Cache County Dementia Progression Study. *Journal of Gerontology: Psychological Sciences Sci Soc Sci, 64B*(5), 560-568. https://doi.org/10.1093/geronb/gbp052 (Journals of Gerontology: Series B)

Nowrangi, M. A., Lyketsos, C. G., & Rosenberg, P. B. (2015). Principles and management of neuropsychiatric symptoms in Alzheimer's dementia. *Alzheimer's Research & Therapy, 7*(1), 12-12. https://doi.org/10.1186/s13195-015-0096-3

O'Rourke, H. M., Duggleby, W., Fraser, K. D., & Jerke, L. (2015). Factors that affect quality of life from the perspective of people with dementia: A metasynthesis. *Journal of American Geriatric Society, 63*(1), 24-38. https://doi.org/10.1111/jgs.13178

O'Rourke, H. M., Fraser, K. D., & Duggleby, W. (2015). Does the quality of life construct as illustrated in quantitative measurement tools reflect the perspective of people with dementia? *Journal of Advanced Nursing, 71*(8), 1812-1824. https://doi.org/10.1111/jan.12667

Onwuegbuzie, A. J., & Johnson, B. R. (2006). The Validity Issue in Mixed Research. *Research in the schools, 13*(1), 48.

Osman, S. E., Tischler, V., & Schneider, J. (2016). 'Singing for the Brain': A qualitative study exploring the health and well-being benefits of singing for people with dementia and their carers. *Dementia: The International Journal of Social Research and Practice, 15*(6), 1326-1339. https://doi.org/10.1177/1471301214556291

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of All Pairs. *Behavior Therapy, 40*(4), 357-367. https://doi.org/10.1016/j.beth.2008.10.006

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification, 35*(4), 303-322. https://doi.org/10.1177/0145445511399147

Perrin, T. (1997). The Positive Responses Schedule for severe dementia. *Aging & Mental Health, 1*, 184–191. https://doi.org/10.1080/13607869757290

Perugia, G., van Berkel, R., Díaz-Boladeras, M., Català-Mallofré, A., Rauterberg, M., & Barakova, E. (2018). Understanding engagement in dementia through behavior. The ethographic and Laban-inspired coding system of engagement (ELICSE) and the Evidence-based Model of Engagement-related Behavior (EMODEB). *Frontiers in Psychology, 9*, 1-18. https://doi.org/10.3389/fpsyg.2018.00690

Petty, S., Harvey, K., Griffiths, A., Coleston, D. M., & Dening, T. (2018). Emotional distress with dementia: A systematic review using corpus-based analysis and meta-ethnography. *International Journal Geriatric Psychiatry, 33*(5), 679-687. https://doi.org/10.1002/gps.4870

Piccirillo, M. L., Beck, E. D., & Rodebaugh, T. L. (2019). A clinician's primer for idiographic research: Considerations and recommendations. *Behavior Therapy, 50*(5), 938-951. https://doi.org/10.1016/j.beth.2019.02.002

Prince, M., Ali, G.-C., Guerchet, M., Prina, A. M., Albanese, E., & Wu, Y.-T. (2016). Recent global trends in the prevalence and incidence of dementia, and survival with dementia. *Alzheimer's Research & Therapy, 8*(1), 23-23. https://doi.org/10.1186/s13195-016-0188-8

Prince, M., Wimo A, Guerchet M, Ali G. C, Wu Y. T, & M., P. (2015). *World Alzheimer report 2015 - the global impact of dementia: an analysis of prevalence, incidence, cost and trends*. Alzheimer's Disease International.

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147-1157. https://doi.org/10.1007/s11136-018-1798-3

Pustejovsky, J. E. (2015). Measurement-comparable effect sizes for single-case studies of free-operant behavior. *Psychological methods, 20*(3), 342-359. https://doi.org/10.1037/met0000019

Pustejovsky, J. E. (2018). Using response ratios for meta-analyzing single-case designs with behavioral outcomes. *Journal of School Psychology, 68*, 99-112. https://doi.org/10.1016/j.jsp.2018.02.003

Pustejovsky, J. E. (2019). Procedural sensitivities of effect sizes for single-case designs with directly observed behavioral outcome measures. *Psychological methods, 24*(2), 217-235. https://doi.org/10.1037/met0000179

Pustejovsky, J. E., & Swan, D. M. (2018). *Single-case effect size calculatur (Version 0.5) Web application.* In

Pustejovsky, J. E., Swan, D. M., & English, K. W. (2019). An examination of measurement procedures and characteristics of baseline outcome data in single-case research. *Behavioral Modification*, 1-32. https://doi.org/10.1177/0145445519864264

R Core Team. (2020). *R: A language and environment for statistical computing.* In R Foundation for Statistical Computing. https://www.R-project.org/.

Raglio, A., Bellelli, G., Traficante, D., Gianotti, M., Ubezio, M. C., Villani, D., & Trabucchi, M. (2008). Efficacy of music therapy in the treatment of behavioral and psychiatric symptoms of dementia. *Alzheimer Disease and Associated Disorders, 22*(2), 158-162. https://doi.org/10.1097/WAD.0b013e3181630b6f

Reid, A.-M., Brown, J. M., Smith, J. M., Cope, A. C., & Jamieson, S. (2018). Ethical dilemmas and reflexivity in qualitative research. *Perspectives of Medical Education, 7*(2), 69-75. https://doi.org/10.1007/s40037-018-0412-2

Reilly, S. T., Harding, A. J. E., Morbey, H., Ahmed, F., Williamson, P. R., Swarbrick, C., Leroi, I., Davies, L., Reeves, D., Holland, F., Hann, M., & Keady, J. (2020). What is important to people with dementia living at home? A set of core outcome items for use in the evaluation of non-pharmacological community-based health and social care interventions. *Age and Ageing, 49*(4), 664-671. https://doi.org/10.1093/ageing/afaa015

Rolvsjord, R., Gold, C., & Stige, B. (2005). Research rigour and therapeutic flexibility: Rationale for a therapy manual developed for a randomised controlled trial. *Nordic Journal of Music Therapy, 14*(1), 15-32. https://doi.org/10.1080/08098130509478122

RStudio Team. (2020). *RStudio: Integrated Development for R.* RStudio. http://www.rstudio.com/.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of Psychological Well-Being. *Journal of Personality and Social Psychology, 57*(6), 1069-1081. https://doi.org/10.1037/0022-3514.57.6.1069

Santana-Berlanga, N. d. R., Porcel-Gálvez, A. M., Botello-Hermosa, A., & Barrientos-Trigo, S. (2020, 2020/03/14/). Instruments to measure quality of life in institutionalised older adults: Systematic review. *Geriatric Nursing*. https://doi.org/https://doi.org/10.1016/j.gerinurse.2020.01.018

Sauer, P. E., Fopma-Loy, J., Kinney, J. M., & Lokon, E. (2016). "It makes me feel like myself": Person-centered versus traditional visual arts activities for people with dementia [Art & Music & Movement Therapy 3357]. *Dementia: The International Journal of Social Research and Practice, 15*(5), 895-912. https://doi.org/http://dx.doi.org/10.1177/1471301214543958

Schall, A., Haberstroh, J., & Pantel, J. (2015). Time series analysis of individual music therapy in dementia. Effects on communication behavior and emotional well-being. *GeroPsych, 28*(3), 113-122. https://doi.org/10.1024/1662-9647/a000123

Schulz, R., Cook, T. B., Beach, S. R., Lingler, J. H., Martire, L. M., Monin, J. K., & Czaja, S. J. (2013). Magnitude and causes of bias among family caregivers rating Alzheimer's disease patients [Neurological Disorders & Brain Damage 3297]. *The American Journal of Geriatric Psychiatry, 21*(1), 14-25. https://doi.org/http://dx.doi.org/10.1016/j.jagp.2012.10.002

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred Reporting Items for

Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement *BMJ, 349*(Jan 2), g7647. https://doi.org/10.1186/2046-4053-4-1

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology, 4*(1), 1-32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Sihvonen, A. J., Särkämö, T., Leo, V., Tervaniemi, M., Altenmüller, E., & Soinila, S. (2017). Music-based interventions in neurological rehabilitation. *Lancet Neurolology, 16*(8), 648-660. https://doi.org/10.1016/S1474-4422(17)30168-0

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Intepretative Phenomenological Analysis*. Sage Publications.

Smith, M., Gerdner, L. A., Hall, G. R., & Buckwalter, K. C. (2004). History, development, and future of the progressively lowered stress threshold: A conceptual model for dementia care. *Journal of the American Geriatric Society, 52*(10), 1755-1760. https://doi.org/10.1111/j.1532-5415.2004.52473.x

Snyder, M., Ryden, M. B., Shaver, P., Wang, J.-J., Savik, K., Gross, C. R., & Pearson, V. (1998). The Apparent Emotion Rating Instrument: Assessing affect in cognitively impaired elders. *Clinical Gerontologist, 18*(4), 17-29. https://doi.org/10.1300/J018v18n04_03

Spector, A., & Orrell, M. (2010). Using a biopsychosocial model of dementia as a tool to guide clinical practice. *International Psychogeriatrics, 22*(6), 957-965. https://doi.org/10.1017/S1041610210000840

Steinberg, M., Shao, H., Zandi, P., Lyketsos, C. G., Welsh-Bohmer, K. A., Norton, M. C., Breitner, J. C. S., Steffens, D. C., & Tschanz, J. T. (2008). Point and 5-year period prevalence of neuropsychiatric symptoms in dementia: the Cache County Study. *International Journal of Geriatric Psychiatry, 23*(2), 170-177. https://doi.org/10.1002/gps.1858

Steingrimsdottir, H. S., & Arntzen, E. (2015). On the utility of within-participant research design when working with patients with neurocognitive disorders. *Clinical Interventions in Aging, 10*, 1189-1199. https://doi.org/10.2147/CIA.S81868

Stern, R. A., Arruda, J. E., Hooper, C. R., Wolfner, G. D., & Morey, C. E. (1997). Visual analogue mood scales to measure internal mood state in neurologically impaired patients: description and initial validity evidence. *Aphasiology, 11*, 59-71. https://doi.org/10.1080/02687039708248455

Stige, B., Malterud, K., & Midtgarden, T. (2009). Toward an agenda for evaluation of qualitative research. *Qualitative Health Research, 19*(10), 1504-1516. https://doi.org/10.1177/1049732309348501

Stoner, C. R., Stansfeld, J., Orrell, M., & Spector, A. (2019). The development of positive psychology outcome measures and their uses in dementia research: A systematic review. *Dementia: The International Journal of Social Research and Practice, 18*(6), 2085-2106. https://doi.org/10.1177/1471301217740288

Strobel, C., & Engedal, K. (2008). *MMSE-NR. Norsk Revidert Mini Mental Status Evaluering. Revidert og utvidet manual.* Oslo: Nasjonalt kompetansesenter for aldring og helse.

Strohmaier, S., Homans, K. M., Hulbert, S., Crutch, S. J., Brotherhood, E. V., Harding, E., & Camic, P. M. (2021). Arts-based interventions for people living with dementia: Measuring 'in the moment' wellbeing with the Canterbury Wellbeing Scales. *Wellcome open research, 6*, 59. https://doi.org/10.12688/wellcomeopenres.16596.2

Särkämö, T. (2018). Music for the ageing brain: Cognitive, emotional, social, and neural benefits of musical leisure activities in stroke and dementia. *Dementia: The International Journal of Social Research and Practice, 17*(6), 670-685. https://doi.org/10.1177/1471301217729237

Tan, J., Wee, S.-L., Yeo, P. S., Choo, J., Ritholz, M., & Yap, P. (2019). A new music therapy engagement scale for persons with dementia [Tests & Testing 2220]. *International Psychogeriatrics, 31*(1), 49-58. https://doi.org/http://dx.doi.org/10.1017/S1041610218000509

Tarlow, K. R. (2016). An improved rank correlation effect size statistic for single-case designs: Baseline Corrected Tau. *Behavior Modification, 41*(4), 427-467. https://doi.org/10.1177/0145445516676750

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., Horner, R., Kazdin, A., Kratochwill, T., McDonald, S., Sampson, M., Shamseer, L., Togher, L., Albin, R., Backman, C., Douglas, J., Evans, J. J., Gast, D., Manolov, R., Mitchell, G., Nickels, L., Nikles, J., Ownsworth, T., Rose, M., Schmid, C. H., & Wilson, B. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 statement. *Journal of Clinical Epidemiology, 73*, 142-152. https://doi.org/10.1016/j.jclinepi.2016.04.006

Teddlie, C. (2010). Overview of contemporary issues in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (2 ed., pp. 1-51). SAGE Publications, Inc. https://doi.org/10.4135/9781506335193.n1

Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 18*(8), 1115. https://doi.org/10.1007/s11136-009-9528-5

Terwee, C. B., Prinsen, C., Chiarotto, A., Westerman, M., Patrick, D., Alonso, J., Bouter, L., Vet, H., & Mokkink, L. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research, 27*(5), 1159-1170. https://doi.org/10.1007/s11136-018-1829-0

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of educational and behavioral statistics, 40*(6), 604-634. https://doi.org/10.3102/1076998615606099

Tschanz, J., Hess, K., Corcoran, C., Snyder, C., Wanzek, J., Norton, M., Piercy, K., Rabins, P., Fauth, E., & Lyketsos, C. (2012). Caregiver coping strategies predict neuropsychiatric symptoms in dementia: The Cache County Dementia Progression Study. *Alzheimer's & dementia, 8*(4), P126-P126. https://doi.org/10.1016/j.jalz.2012.05.332

van der Steen, J. T., Smaling, H. J., van der Wouden, J. C., Bruinsma, M. S., Scholten, R. J., & Vink, A. C. (2018). Music-based therapeutic interventions for people with dementia. *Cochrane Database of Systematic Reviews*(7). https://doi.org/10.1002/14651858.CD003477.pub4.

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of counseling and development, 93*(4), 403-411. https://doi.org/10.1002/jcad.12038

Vogelpohl, T. S., & Beck, C. K. (1997). Affective responses to behavioral interventions. *Seminars in Clinical Neuropsychiatry, 2*(2), 102-112. https://doi.org/10.1053/SCNP00200102

Vogt, D. S., King, D. W., & King, L. A. (2004). Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. *Psychological Assessment, 16*(3), 231-243. https://doi.org/10.1037/1040-3590.16.3.231

Von Kutzleben, M., Schmid, W., Halek, M., Holle, B., & Bartholomeyczik, S. (2012). Community-dwelling persons with dementia: What do they need? What do they demand? What do they do? A systematic review on the subjective experiences of persons with dementia. *Aging & Mental Health, 16*(3), 378-390. https://doi.org/10.1080/13607863.2011.614594

Wadham, O., Simpson, J., Rust, J., & Murray, C. (2016). Couples' shared experiences of dementia: a meta-synthesis of the impact upon relationships and couplehood. *Aging Ment Health, 20*(5), 463-473. https://doi.org/10.1080/13607863.2015.1023769

Warner, J., McCarney, R., Griffin, M., Hill, K., & Fisher, P. (2008). Participation in dementia research: rates and correlates of capacity to give informed consent.

Journal of Medical Ethics, 34(3), 167-170. https://doi.org/10.1136/jme.2006.019786

Whittemore, R., Chase, S. K., & Mandle, C. L. (2001, 2001/07/01). Validity in Qualitative Research. *Qualitative Health Research, 11*(4), 522-537. https://doi.org/10.1177/104973201129119299

Wiggins, B. J. (2011, Feb). Confronting the dilemma of mixed methods. *Journal of Theoretical and Philosophical Psychology, 31*(1), 44-60. https://doi.org/10.1037/a0022612

Williams, C. L., Newman, D., & Hammar, L. M. (2017). Preliminary psychometric properties of the verbal and nonverbal interaction scale: An observational measure for communication in persons with dementia. *Issues in Mental Health Nursing, 38*, 381-390. https://doi.org/10.1080/01612840.2017.1279248

Winblad, B. P., Amouyel, P. P., Andrieu, S. P., Ballard, C. P., Brayne, C. P., Brodaty, H. P., Cedazo-Minguez, A. P., Dubois, B. P., Edvardsson, D. P., Feldman, H. P., Fratiglioni, L. P., Frisoni, G. B. P., Gauthier, S. P., Georges, J. B. A., Graff, C. P., Iqbal, K. P., Jessen, F. P., Johansson, G., Jönsson, L. M. D., … Zetterberg, H. P. (2016). Defeating Alzheimer's disease and other dementias: a priority for European science and society. *Lancet Neurology, 15*(5), 455-532. https://doi.org/10.1016/S1474-4422(16)00062-4

Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in quantitative methods for psychology, 8*(1), 52-69. https://doi.org/10.20982/tqmp.08.1.p052

Wood, W. (2005). Toward developing new occupational science measures: An example from dementia care research. *Journal of occupational science, 12*(3), 121-129. https://doi.org/10.1080/14427591.2005.9686555

Woods, B. (1999). Promoting well-being and independence for people with dementia. *International Journal of Geriatric Psychiatry, 14*(2), 97-105. https://doi.org/10.1002/(sici)1099-1166(199902)14:2<97::aid-gps941>3.0.co;2-x

Woods, B. (2012). Well-being and dementia – how can it be achieved? *Quality in Ageing: Policy, Practice and Research, 13*(3), 205-211. https://doi.org/10.1108/14717791211264214

World Health Organisation. (2016). *ICD-10: International statistical classification of diseases and related health problems:* (5th ed.). World Health Organization.

World Health Organization. (2017). *Global action plan on the public health response to dementia 2017–2025*. Geneva: World Health Organisation

World Medical Association (2013) Declaration of Helsinki. Ethical principles for medical research involving human subjects, 1-4. https://apps.who.int/iris/handle/10665/268312

Wu, Y.-T., Beiser, A. S., Breteler, M. M. B., Fratiglioni, L., Helmer, C., Hendrie, H. C., Honda, H., Ikram, M. A., Langa, K. M., Lobo, A., Matthews, F. E., Ohara, T., Pérès, K., Qiu, C., Seshadri, S., Sjölund, B.-M., Skoog, I., & Brayne, C. (2017). The changing prevalence and incidence of dementia over time - current evidence. *Nature Reviews Neurology, 13*(6), 327-339. https://doi.org/10.1038/nrneurol.2017.63

Ydstebø, A. E., Benth, J. Š., Bergh, S., Selbæk, G., & Vossius, C. (2020). Informal and formal care among persons with dementia immediately before nursing home admission. *BMC Geriatrics, 20*(1), 296-296. https://doi.org/10.1186/s12877-020-01703-8

Young, Y., Frick, K. D., & Phelan, E. A. (2009). Can successful aging and chronic illness coexist in the same individual? A multidimensional concept of successful aging. *Journal of Americal Medical Directors Association (JAMDA), 10*(2), 87-92. https://doi.org/10.1016/j.jamda.2008.11.003

Zhang, Y., Cai, J., An, L., Hui, F., Ren, T., Ma, H., & Zhao, Q. (2017). Does music therapy enhance behavioral and cognitive function in elderly dementia patients? A systematic review and meta-analysis. *Ageing Research Reviews, 35*, 1-11. https://doi.org/10.1016/j.arr.2016.12.003

Øksnebjerg, L., Diaz-Ponce, A., Gove, D., Moniz-Cook, E., Mountain, G., Chattat, R., & Woods, B. (2018). Towards capturing meaningful outcomes for people with dementia in psychosocial intervention research: A pan-European consultation. *Health Expectations, 21*(6), 1056-1065. https://doi.org/10.1111/hex.12799

Appendices

Appendix 1 -Paper 1: Adapted RoB-ratings for COSMIN Box 1, 2 and 3

	Adapted Cosmin guidelines for evaluating the content validity of observer-reported outcome instruments (ObsROMS)[1]
	Box 1 – ObsROM development
1a	Instrument design
1	Is a clear description provided of the construct to be measured?
2	Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured?
3	Is a clear description provided of the target population for which the ***ObsROM*** was developed?
4	Is a clear description provided of the context of use (i.e. discriminative, evaluative purpose, and/or predictive)
5	Was the ***ObsROM*** development study performed in a sample representing the target population for which the instrument was developed?
	Concept elicitation
6	Was an appropriate ***qualitative or quantitative*** data collection method used to identify relevant ***items*** for a new ***ObsROM***[2]?
7	***If qualitative:*** Were skilled group moderators/interviewers/ ***observators/ raters*** used?
8	***If qualitative interviews:*** Were the group meetings or interviews based on an appropriate topic or interview guide, interviews were recorded and transcribed verbatim?
9	***If qualitative***: Was at least part of the data coded independently? (11-49% = A)
10	***If qualitative:*** Was data collection/fieldwork continued until saturation was reached?
11	***If theoretical approach: Was a literature review or conceptual framework clearly described and a thorough representation of the research field?***
12	For quantitative studies (i.e. survey or Delphi procedure): Was the data collection continued until consensus was reached?
13	For quantitative studies ***(i.e. survey or Delphi procedure):*** was the sample size appropriate?
14	***If qualitative fieldwork: Was a methodological sound approach to develop coding scheme or ethogram applied***
15	***Were at least two appropriate approaches to developing the instrument used (qualitative fieldwork using observation, qualitative interviews/focus groups, theoretical approach building on former instruments or theoretical***

---

[1] The adapted guidelines are based on the COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures (Mokkink et al., 2017; Terwee et al., 2018) which is available for download on. and specific criteria for establishing content validity of observational instruments based on recommendations from Bakeman and Quera (2011) and Chorney, McMurtry, Chambers, and Bakeman (2015). Changes from the original checklist is marked with ***bold and italic writing.***

[2] Finding relevant items is based on identifying issues important to the patients and extracting relevant items from these issues (Brod, Tesler, & Christensen, 2009). For observational instruments we recognize several approaches appropriate for identifying relevant issues. The main criteria is to show adequate research rigour in the specific research approach

	development, OR quantitative survey/Delphi approach with relevant experts (patients, family-caregivers, or clinicians)
16	Was an appropriate ***and transparent*** approach used to analyze the data?
1b	Pilot test/development phase
17	Was pilot test performed?
	General design requirements
18	Was ***pilot testing*** performed in a sample representing the target population ***and study context***?
19	***Were operational definitions refined based on pilot-testing?***
20	Was each item tested in an appropriate number of patients? For qualitative studies ≥7 For quantitative (survey) studies ≥50
	Comprehensibility
21	Were problems regarding the comprehensibility of the instrument instructions, items, response options, and recall period appropriately addressed by adapting the instrument?
22	Were all the items tested in their final form?
	Box 2. Content validity[3]
	Asking patients ***or family caregivers*** about relevance[4]
1	Was an appropriate method used to ask patient or ***family caregivers*** whether each item is relevant for their experience with the condition?
2	Was each item tested in an appropriate number of patients? For qualitative studies ≥7 For quantitative (survey) studies ≥50
3	***If qualitative:*** Were skilled group moderators/interviewers used?
4	***If qualitative interviews***: Were the group meetings or interviews based on an appropriate topic or interview guide, interviews were recorded and transcribed verbatim?
5	Was an appropriate approach used to analyses the data?
6	***If qualitative***: Were at least two researchers involved in the analysis?
7	***For quantitative studies (i.e. survey or Delphi procedure): Was the data collection continued until consensus about relevance was reached?***
	Asking patients ***or family caregivers*** about comprehensiveness
8	Was an appropriate method used to for assessing the comprehensiveness of the ***ObsROM***
9	Was each item tested in an appropriate number of patients? For qualitative studies ≥7 For quantitative (survey) studies ≥50
10	***If qualitative:*** Were skilled group moderators/interviewers used?
11	***If qualitative interviews***: Were the group meetings or interviews based on an appropriate topic or interview guide, interviews were recorded and transcribed verbatim?

---

[3] The difference between development studies and studies on content validity is that the latter refers to studies conducted after the final ObsROM-version. Criteria is that a new sample of participants are included (independent of the development study)

[4] For PROMS, patients should be asked about relevance, comprehensiveness and comprehensibility of the instrument. For observational instruments, inclusion of patients is an advantage, but relevance, comprehensiveness and comprehensibility need to be evaluated at minimum by the observers actually scoring the instrument.

12	Was an appropriate approach used to analyze the data?
13	***If qualitative***: Were at least two researchers involved in the analysis?
14	***For quantitative studies (i.e. survey or Delphi procedure): Was the data collection continued until consensus about comprehensiveness was reached?***
	Asking professionals about relevance, ***or investigating relevance with other approach***
15	***Was the relevance of items secured by an appropriate method? (Fieldwork/ethogram, adapting other coding schemes, theoretical approach, literature search, qualitative interviews or quantitative surveys involving professionals?)***
16	***If survey or qualitative interview***: Were professionals from all relevant disciplines included?
17	Was each item tested in an appropriate number of professionals ***or subjects***? For qualitative studies ≥7 For quantitative (survey) studies ≥50
18	***If qualitative approach:*** Were skilled group moderators/interviewers /raters/ observers used?
19	**If qualitative interviews:** Were the group meetings or interviews based on an appropriate topic or interview guide, interviews were recorded and transcribed verbatim?
20	***If qualitative***: Were at least two researchers involved in the analysis?
21	Was an appropriate approach used to analyze the data?
22	***For quantitative studies (i.e. survey or Delphi procedure): Was the data collection continued until consensus about relevance was reached?***
	Asking professionals about the comprehensiveness, ***or investigating comprehensiveness with other approach***
23	***Was an appropriate method used for assessing the comprehensiveness of the instrument (Fieldwork/ethogram, adapting other coding schemes, theoretical approach, literature search, qualitative interviews or quantitative surveys involving professionals?)***
24	***If survey or qualitative interview***: Were professionals from all relevant disciplines included?
25	Was each item tested in an appropriate number of professionals or ***subjects***? For qualitative studies ≥7 For quantitative (survey) studies ≥50
26	Was an appropriate approach used to analyze the data?
27	***If qualitative: Were skilled group moderators/interviewers/raters/ observers used?***
28	***If qualitative interviews: Were the group meetings or interviews based on an appropriate topic or interview guide, interviews were recorded and transcribed verbatim?***
29	***If qualitative:*** Were at least two researchers involved in the analysis?
30	***For quantitative studies (i.e. survey or Delphi procedure): Was the data collection continued until consensus about relevance was reached?***

Appendix 2 – Paper 1: Supplementary Table S1 Measurement properties and study characteristics and Supplementary Table S2 Feasibility and interpretability

Available at the data depository of UiB from the following link: Vedlegg (uib.no)

Appendix 3 – Paper 3: Interview guide for focus groups

Interview-guide

This focus group aims at exploring the understanding of well-being expressions in people with dementia. In particular, we aim to investigate which expressions you think is important when observing a person during [psychosocial interventions/music therapy].

Part 1:

How do you understand the term well-being? How is well-being expressed?

How do you understand the term well-being in dementia? How is it expressed?

How can you observe well-being in people living with dementia? Think about persons with mild, moderate, and severe dementia. Feel free to provide specific and anonymous examples from persons you are familiar with.

Which nonverbal signs do you think is important to interpret presence of well-being in people living with dementia in a [psychosocial/music therapeutic] setting (i.e., milieu therapy or activities)? Feel free to provide specific and anonymous examples from persons you are familiar with.

Which verbal signs do you think is important to interpret presence of well-being in people living with dementia in a [psychosocial/music therapeutic] setting (i.e., milieu therapy or activities)? Feel free to provide specific and anonymous examples from persons you are familiar with.

Part 2 [Handing out OWLS]

This instrument is created to observe well-being in the moment for people living with dementia. This scale is designed to be used for people living with dementia throughout the different phases. It is built on theory about components of well-being in dementia, i.e., Lawton and Kitwood. In the literature, well-being is defined as presence of positive affect and absence of negative affect. Meanwhile, it has been complicated to measure negative affect, and negative expressions are completely omitted in this scale.

The scale utilizes video-recordings of people during an intervention aiming to facilitate well-being. It could be milieu therapy, activities, music therapy, reminiscence therapy and other psychosocial interventions for people living with dementia. One problem with existing observational scales is that many focus on verbal expressions of well-being. During observation of creative interventions, such as music therapy, the score will be very low if the person is playing an instrument and is not making verbal expressions. Use some minutes to read the instructions for the scales, the operationalizations of the different items that is to be observed, and imagine you are observing a person in 30 seconds and scoring the different items. How comprehensive is the wording in the scale? What is unclear? Is any information missing?

How relevant does the 10 items appear for observing people with dementia?

How relevant does the 10 items appear for people with dementia in [a psychosocial setting/music therapy]?

How relevant does the operationalization (the explanation of the behaviors and expressions for each item) appear for observing people with dementia? How relevant does the operationalization appear for observing people with dementia in [a psychosocial setting/music therapy]?

How comprehensive is the different verbal and nonverbal expressions of well-being is the scale? Are any elements missing to fully cover well-being in people living with dementia? Be as specific as possible when considering what is missing to cover observation of well-being in [a psychosocial setting/music therapy]

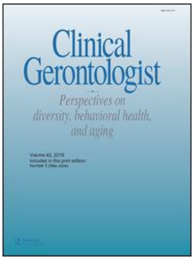
Are any items clearly describing something else than well-being? Please be as specific as possible.

How do you think it would be to use this scale?

Paper 2:

Madsø, K. G., Molde, H, Hynninen, K. M., & Nordhus, I. H. (2021). Observing music therapy in dementia: Repeated single-case studies assessing well-being and sociable interaction. *Clinical Gerontologist*. Advance online publication. https://doi.org/10.1080/07317115.2021.1978121

Paper 2



Clinical Gerontologist

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/wcli20

Observing Music Therapy in Dementia: Repeated Single-case Studies Assessing Well-being and Sociable Interaction

Kristine Gustavsen Madsø, Helge Molde, Kia Minna Hynninen & Inger Hilde Nordhus

To cite this article: Kristine Gustavsen Madsø, Helge Molde, Kia Minna Hynninen & Inger Hilde Nordhus (2021): Observing Music Therapy in Dementia: Repeated Single-case Studies Assessing Well-being and Sociable Interaction, Clinical Gerontologist, DOI: 10.1080/07317115.2021.1978121

To link to this article: https://doi.org/10.1080/07317115.2021.1978121

CLINICAL GERONTOLOGIST
https://doi.org/10.1080/07317115.2021.1978121

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Observing Music Therapy in Dementia: Repeated Single-case Studies Assessing Well-being and Sociable Interaction

Kristine Gustavsen Madsø CandPsychol[a,b], Helge Molde PhD[a], Kia Minna Hynninen PhD[b], and Inger Hilde Nordhus PhD[a,c]

[a]Department of Clinical Psychology, University of Bergen, Bergen, Norway; [b]NKS Olaviken Gerontopsychiatric Hospital, Bergen, Askøy, Norway; [c]Department of Behavioral Medicine, University of Oslo, Oslo, Norway

ABSTRACT

Objectives: This study compared behavioral expressions of momentary well-being and sociable behavior toward significant others during music therapy and regular social interaction.
Methods: A 10-week active music therapy intervention was provided for people living with dementia and family caregivers. A bi-phasic AB single-case design was replicated for three sessions per dyad and coded using the Observable Well-being in Living with Dementia-Scale (OWLS) and the Verbal and Nonverbal Sociable Interaction Scale-Care Receiver (VNVIS-CR). Effect sizes (Log Response Ratio) were calculated for each session and analyzed with robust cluster meta-analysis.
Results: Eleven dyads were included, and 32 sessions analyzed (2102 observations). Within sessions we found a 48% increase in well-being, and a 32% increase in sociable interaction during music therapy. Heterogeneity was high. Dementia severity predicted an increase in nonverbal sociable interaction (93% for moderate dementia). Depression and time did not predict any change.
Conclusion: The potential of music therapy to increase well-being and sociable interactions toward significant others calls for further investigation of heterogeneity and covariates. Single-case designs are demonstrated to be feasible for these investigations.
Clinical implications: Preference-based music therapy may alleviate some of the individual and relational consequences of living with dementia, facilitating positive emotions and connection to significant others.

KEYWORDS
Music therapy; dementia; observation; single-case design; well-being; sociable interaction; caregiver

Introduction

An increasing attention toward living well with dementia is present in dementia research and care (Dröes et al., 2017), public action plans, and guidelines (i.e. National Institute for Healt and Care Excellence, 2018; World Health Organization [WHO], 2017). Key targets for psychosocial interventions emphasized by home-dwelling people living with dementia are coping with psychological distress following the diagnosis, maintaining normality and identity, participating in meaningful and enjoyable activities, and having good social relationships (Miranda-Castillo, Woods, & Orrell, 2013; Øksnebjerg et al., 2018; Reilly et al., 2020; von Kutzleben, Schmid, Halek, Holle, & Bartholomeyczik, 2012).

Personalized and active music interventions may be beneficial for several of these objectives for multiple reasons (Brancatisano, Baird, & Thompson, 2020). Music is a strong trigger of positive emotions (e. g., Juslin, 2013), and familiar music may trigger pleasant memories and maintain a sense of identity and coherence in the individual living with dementia (Baird & Thompson, 2018; Särkämö, 2018). Additionally, music is an engaging and inherently social activity (Brancatisano et al., 2020). Furthermore, active music interventions may positively affect cognition (Fusar-Poli, Bieleninik, Brondino, Chen, & Gold, 2018), and meet current recommendations of individualizing interventions for this group (Dowson, McDermott, & Schneider, 2019; WHO, 2017).

CONTACT Kristine Gustavsen Madsø kristine.madso@uib.no Department of Clinical Psychology, University of Bergen, PB 7807, 5020 Bergen, Norway

Supplemental data for this article can be accessed on the publisher's website.

While meta-analyses of music intervention studies have shown small significant effects on emotional well-being at end of treatment (van der Steen et al., 2018; Zhang et al., 2017), most randomized controlled trials have been conducted in long-term care or hospital settings (Cho, 2018; Cooke, Moyle, Shum, Harrison, & Murfield, 2010; Hsu, Flowerdew, Parker, Fachner, & Odell-Miller, 2015; Raglio et al., 2015; Ridder, Stige, Qvale, & Gold, 2013). However, most people living with dementia are cared for in their homes by family caregivers (Livingston et al., 2017).

Family caregivers report that some of the most distressing aspects of caregiving are the disruptive behaviors commonly described as the behavioral and psychological symptoms of dementia or BPSD (Cheng, 2017; Kales, Gitlin, & Lyketsos, 2015). Most people living with dementia experience these symptoms, with apathy as the most frequent (Livingston et al., 2017). While BPSD are driven by a diversity of causes clustered in the person, caregiver and/or environment (Kales et al., 2015), recent meta-analyses conclude that music therapy may decrease BPSD symptoms (Abraha et al., 2017; van der Steen et al., 2018). Music therapy may be beneficial for both caregivers and care receivers. Thus, including dyads in music therapy interventions seems both clinically relevant and feasible.

To our knowledge, only one randomized controlled trial has included family caregivers in music interventions. Särkämö et al. (2014) found that group singing interventions improved mood in home-dwelling people living with dementia, and decreased family caregiver burden. Other small-scale studies have shown musical interventions to positively affect the relationship of the dyads in qualitative (Baker, Grocke, & Pachana, 2012; Camic, Williams, & Meeten, 2013; Clark, Tamplin, & Baker, 2018; Davidson & Fedele, 2011), as well as quantitative measures (Clair, 2002).

Changes in social behavior is common in dementia (Livingston et al., 2017), and it is reasonable to expect interventions targeting social behavior to be of value to the caregiver and care-receiver dyad. Documentation of the effect of music therapy on sociable behavior in dementia is sparse (van der Steen et al., 2018), but single-case studies have shown increased communication behavior (Schall, Haberstroh, & Pantel, 2015), and mutual engagement (Clair, 2002).

Most clinical studies assessing music therapy administer scales before and after the intervention period (van der Steen et al., 2018). However, the degenerative nature of dementia may mask clinically relevant changes over shorter time periods and does not necessarily reflect a lack of treatment effects. To assess clinically and socially meaningful outcomes for people living with dementia, increasing the sensitivity of measurement instruments and research designs are recommended (Cho, 2018; Dowson et al., 2019; Schall et al., 2015). Single-case designs may be a feasible approach for smaller samples and are underused in dementia research (Steingrimsdottir & Arntzen, 2015). These designs have a high ecological validity and provide explanatory power when the person serves as their control, the intervention is replicated, and threats to internal and external validity are managed (Manolov, Gast, Perdices, & Evans, 2014).

Aims

Thus, the goal of this project was to investigate the effect of individually tailored music therapy for home-dwelling people living with dementia involving close family caregivers as a collateral therapist. The primary focus was the potential effect on well-being in the care receiver, and sociable behavior toward the caregiver. Additionally, we aimed to demonstrate the utility of a single-case design using fine-grained outcome measures.

Our primary hypothesis was that individually tailored music therapy would increase observed within-person expressions of momentary well-being and sociable behavior when compared to a *baseline* of regular social interaction, *within* and *across sessions*. Our secondary hypotheses were positive changes in self-rated emotions when comparing *pre-* and *post-session* measures, and stable neuropsychiatric symptoms, increased self-reported long-term well-being, and decreased caregiver burden from *pre-* to *post-intervention.*

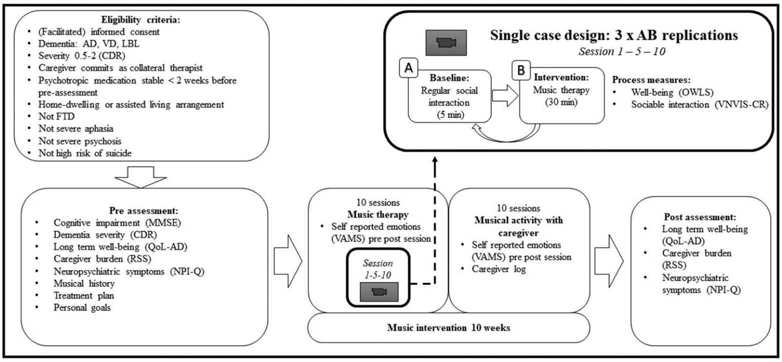


Figure 1. Study design and measurement. *Abbreviations* AD = Alzheimer's dementia, VD = Vascular dementia, LWBD = Dementia with Lewy Bodies, FTLD = Frontotemporal dementia. CDV = Clinical Dementia Rating, MMSE = Mini Mental State Examination, CDV = Clinical Dementia Rating, QoL-AD = Quality of Life in Alzheimer Dementia, RSS = Relative Stress Scale, NPI-Q = Neuropsychiatric Inventory-Questionnaire, VAMS = Visual Analog Mood Scale, OWLS = Observable Well-being in Living with dementia-Scale, VNVIS-CR = Verbal and Nonverbal Sociable Interaction Scale-Care Receiver

Methods

Recruitment

Participants were recruited from the outpatient clinic of NKS Olaviken Gerontopsychiatric Hospital in Bergen, Norway. Eligibility criteria are displayed in Figure 1. The severity of dementia was staged with the Clinical Dementia Rating (CDR; Hughes, Berg, Danziger, Coben, & Martin, 1982), and the Mini-Mental State Examination – Norwegian Revision (Folstein, Folstein, & McHugh, 1975; Strobel & Engedal, 2008).

Ethics

The study protocol was pre-registered (www.clinicaltrials.gov, ID: NCT03011723), and approved by the Regional Committees for Medical and Health Research Ethics in Norway (2016/1374). People eligible for inclusion were invited to participate during a regular appointment at the outpatient clinic and provided with written information about the project. Participants were given time to discuss with a person they trusted before actively registering their interest. Next, an appointment was scheduled in the participants' home, where information about the project was given in person. Informed consent was signed immediately after informing them about the project, facilitating the ability to remember the details about the project and actively deciding if they wanted to participate. Thus, all participants could provide informed consent. Both researchers and music therapists were attentive to signs of withdrawn consent during the study, and repetition of information about the research project was provided when needed (cfr. Dewing, 2007).

Intervention

The active music intervention aimed for two weekly sessions over 10 weeks. Each week, a professional music therapist came to the participant's home for the first weekly session with the dyad. This intervention followed a manual of resource-oriented principles for music therapy (Rolvsjord, Gold, & Stige, 2005). The music therapist guided the dyads in choosing musical activities for the second weekly session, initiated by the collateral therapist. The main principles and content of the music therapy is summarized in Table 1.

Table 1. Music therapy intervention.

Main principles of resource-oriented music therapy[a]	Musical activities with therapist or caregiver
Setting the goals of the therapy together with the client	Singing together (often with music therapist playing guitar, piano, or accordion)
Attentive to the strengths, potentials, and competence of the client	Playing instrumental music together (for example drumming, guitar or harmonica)
Focusing on the musical identity and musical history of the client	Improvising music together
Facilitating positive emotions, emotional engagement, and emotion regulation through music	Listening to live or recorded music
Using music to foster self-perception	Moving to music alone or together
Using music to foster social relationships and communication	Relaxation exercises to music
Session length is approximately 45 minutes, always tailored to the specific needs, attention span, or energy level of the person with dementia	The abovementioned activities were often followed by conversations about music and/or memories coming to mind

[a]These principles originate from the manual of Rolvsjord et al. (2005)

Procedures

The study design is presented in Figure 1. First, the personal musical history of the person with dementia and the shared musical history of the dyad were mapped. A treatment plan including personal goals was made in collaboration with the dyad.

Before and after the 10-week intervention period, long-term well-being, caregiver burden, and neuropsychiatric symptoms were assessed. For each music therapy and collateral led-session, self-reported emotional state before and after the session was measured. Music therapists and collaterals logged the activities in each session. The collaterals rate their session as “negative,” “neutral” or “positive,” to detect adverse effects.

The first, fifth, and tenth music therapist-led sessions were video recorded. Music therapists were instructed to ensure at least five minutes of regular social interaction between the dyad and music therapist, constituting a baseline-phase (A). The following 30 minutes of music therapy with the dyad constituted the intervention-phaseFor each 30 second interval, we dichotomously scored the presence or absence of each behavior (described under “Measures”), using the software Noldus Observer XT 12.5© (Noldus Information Technology, 2015). Three coders including the first author (KGM) were trained until reaching at least 80% inter-rater agreement (Ledford & Gast, 2009). Then, the first 20% of video recordings were coded by main-coder (KGM) and research assistants to assess inter-rater reliability. Feedback to prevent observer drift was given. The assessors were independent of the therapists. No blinding procedure was feasible for the participants, therapists, or coders.

Measures

Primary measures[1]

The Verbal and Nonverbal Interaction Scale – Care Recipient (VNVIS-CR; Williams, Newman, & Hammar, 2017), is a dementia-specific observational instrument consisting of 26 operationalizations of sociable and unsociable verbal and nonverbal behavior toward a close caregiver. A ratio consisting of the number of sociable items divided by the number of unsociable items is calculated per time point (range 0–13). A higher score indicates more sociable behavior. Two subscales exist, sociable-*nonverbal* and sociable-*verbal* interaction. VNVIS-CR is evaluated to have good inter-rater reliability, test–retest reliability, and construct validity (Williams et al., 2017). As VNVIS-CR was not available in Norwegian, formal translation/back-translation was approved by C. L. Williams.

The Observable Well-being in Living with Dementia-Scale (OWLS;[2] Madsø, Pachana & Nordhus; manuscript under review) was developed for this study. The development was based on methodological recommendations (Bakeman & Quera, 2012), including pilot-testing, building on the theory of well-being in dementia, and examining existing observational scales assessing well-being (Algar, Woods, & Windle, 2016; Kitwood, 1997). Finally, the content validity of the scale was established via focus-group-discussions with relevant experts. Construct validity (Terwee et al., 2007), was supported by moderate to strong correlations with total-score and nonverbal score of VNVIS-CR. Responsiveness (Terwee et al., 2007), was supported by moderate correlations between effect sizes and change scores of the Neuropsychiatric Inventory-Questionnaire (NPI-Q, Kaufer et al., 2000), and the self-reported happiness-subscale of the Visual Analog Mood Scale (Stern, Arruda, Hooper, Wolfner, & Morey, 1997). OWLS consists of 10 items ("initiative/response," "attention," "happiness," "enjoyment," "joking," "mastery," "self-confidence," "express identity," "positive feedback," and "relationship"). Two items may need further elaboration. "Express identity" refers to positive initiatives and responses during activity related to personal history or self-perception. "Relationship" refers to initiating turn-taking interactions to achieve closeness with significant others. All operationalizations include both verbal and nonverbal indicators of well-being (range 0–10). Higher scores indicate higher well-being. Scores <2 indicate lack of attention and response toward the activity in the current observations. Items of the scales are presented in Table S2.

Secondary measures

Self-reported emotions were assessed with the Visual Analog Mood Scale (VAMS, Norwegian translation, Stern et al., 1997). Quality of life in Alzheimer's dementia (QoL-AD, Norwegian translation; Logsdon, Gibbons, McCurry, & Teri, 1999) was rated in interviews with people with dementia. Caregivers were interviewed with the Relative Stress Scale (RSS, Norwegian translation; Greene, Gardiner, & Timbury, 1982) and The Neuropsychiatric Inventory-Questionnaire (NPI-Q, Norwegian translation; Kaufer et al., 2000).

Statistical analysis

Statistical analyses were conducted in R (R Core Team, 2020), R studio (RStudio Team, 2020), using the r packages 'ggplot2′ (Wickham, 2016), and "robumeta" (Fisher & Tipton, 2017), and a single-case effect size calculator (Log Response Ratio, LRR, 0.5; Pustejovsky & Swan, 2018).

Single-case analysis for primary measures

For each case, three bi-phasic intra-subject AB-replications were conducted (Tate et al., 2016), to investigate changes in the level of the primary measures. The baseline phase consisted of 10 ratings during regular social interaction, and the intervention phase consisted of 56 ratings during music therapy. The intra-subject replication of the AB-design gave three distinct attempts to investigate the intervention effect, followed by inter-subject replications in 11 cases (Manolov et al., 2014; Tate et al., 2016). No randomization procedure was applied.

The sum of the OWLS-items and the ratio of VNVIS-CR were plotted in R (R Core Team, 2020; Wickham, 2016) and visually inspected (Ledford & Gast, 2009; Tarlow, 2016). As there is no consensus-based method for visual, nonoverlap, or statistical approaches when conducting single-case analyses, we followed the recommendations of Parker, Vannest, and Davis (2011) to visually investigate if there is a) a baseline trend present and b) if there is a strong trend-improvement in the intervention-phase. Different approaches incorporate and control for different parameters, and some single-case effect size estimates are sensitive to study designs such as length of baseline and intervention phase, length of session and type of recording system (Pustejovsky, 2019).

Variability and autocorrelation in time-series are common, and may complicate visual analysis, leading to both over- and underestimation of treatment effects (Parker et al., 2011; Vannest & Ninci, 2015). In these situations, parametric approaches may be helpful. In addition, a baseline trend in the direction of the hypothesized treatment effect may oppose a threat to the internal validity, but correcting for insignificant or random baseline-trends may overcorrect data and mask treatment effects (Tarlow, 2016). To account for the observed characteristics in our data, we chose the parametric approach Log Response Ratio (LRR, Pustejovsky, 2015, 2018). LRR is a promising parametric scale-free approach for calculating single-case effect sizes. The magnitude of the LRR is not sensitive to the sample size and length of the observations, as are several other effect sizes (Pustejovsky, 2019).

Single LRR-estimates do not account for autocorrelation, and variance-estimates may be biased (Pustejovsky, 2015). However, meta-analyses of these effect sizes using robust variance estimation will correct for this (Pustejovsky, 2018). Thus, a meta-analysis of the LRR effect sizes for each individual music therapy session was conducted. Due to the dependency between the effect sizes nested within each case, robust meta-regression is the recommended option (Hedges, 2019). Robust cluster variance estimation with small-sample correction was conducted, with accompanying sensitivity analyses and investigation of heterogeneity with meta-regression (Fisher & Tipton, 2017; Tipton & Pustejovsky, 2015).

Interpreting single-case effect sizes should be based on contextual understanding, and benchmarks may vary (Vannest & Ninci, 2015). The LRR ranges from −1 to +1 but may be converted to "percentage of change," making interpretation straightforward. Pustejovsky (2018) recommends a context-specific interpretation of LRR. In this specific context and based on our knowledge of the specific outcome measures, we interpret a change of 20– 50% to be small, 50–70% to be moderate, and >70% to be large.

Finally, the relative frequencies of the OWLS items and VNVIS-CR nonverbal and verbal interaction items in the baseline and the intervention-phase were explored to investigate the behavioral content of the two phases (total number of occurrences divided by total number of observations).

Statistical analysis for secondary measures

The secondary measures were analyzed with Wilcoxon signed-rank test for paired samples with continuity correction (R Core Team, 2020). Missing data from pre- to post-sessions were omitted, and missing data from pre- to post-intervention were imputed as no change.

Results

A total of 13 participant-dyads were recruited. Two dropped out after one session (withdrawn consent and admission to hospital), and one after 6 sessions (psychotic symptoms). The latter participant's ("Kate") available data was included in the analysis. Of the 11 participants, nine were diagnosed with Alzheimer's dementia and two with Vascular Dementia. They were aged from 71 to 88 years (M =79.82, SD = 5.27), and 63% were women. Clinical dementia stage was very mild for one, mild for five, and moderately severe for five. All participants experienced at least two neuropsychiatric symptoms at inclusion, with symptoms of depression, apathy, and anxiety as the most common. Demographics and clinical characteristics of the 11 participant dyads are presented in Table S1.

Treatment fidelity

Logs showed musical elements were the main ingredient in all the music therapy sessions (range of duration 23– 70 minutes, M = 46.2, SD = 8.85). All participants received 10 music therapy sessions except "Clare" (8 sessions), and "Kate" (6 sessions).

The number of sessions led by the collateral therapist ranged from 0 to 8, with six participants reporting ≥5, and five participants reporting ≤4 sessions (range of duration 10–70 minutes, M = 52, SD = 28.3). Adverse effects were rare. Collateral-sessions were rated in the logs as positive for 86%, neutral for 4%, negative for 4% sessions (6% were not rated).

Single-case analysis

Inter-rater reliability (n = 15 012 ratings) was in the good to excellent range for both instruments, with a mean Cohen's Kappa of 0.92 (CI = 0.87 – .94, p = <0.001) for VNVIS-CR and 0.82 (CI = 0.72–0.89, p = <0.001) for OWLS. Mean inter-rater agreement for VNVIS-CR was 92% (range = 89–93%) and 84% (range = 77–88%) for OWLS.

Plots of all observational data for each single-case are provided in the supplementary material, with an example displayed in Figure 2. Thirty-two sessions were included (N_{baseline} = 320 observations, $N_{\text{intervention}}$ = 1782 observations). Table 2 displays the LRR effect size estimates per session. Phase-specific item-frequency is presented in Table S2.

Investigating validity

An increasing baseline trend was present in one session for well-being ("Ann," Session 1), and three sessions for sociable interaction ("Beth," session 1, "Helen," session 1, and "Kate," session 1). A consistent pattern of returning to baseline between sessions was detected through visual inspection for 30 of 32 observations of

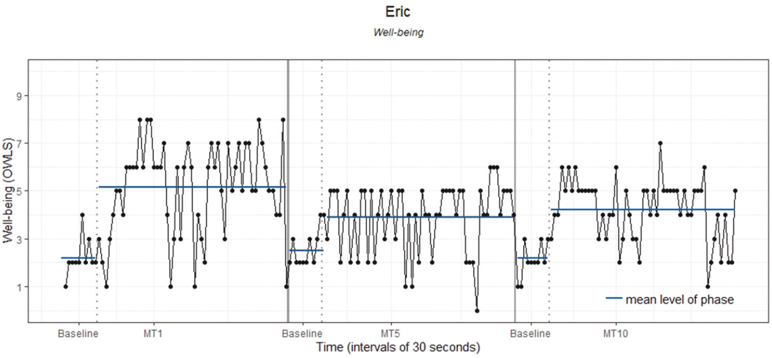


Figure 2. Plot of well-being exemplified by "Eric." *Note*: Plot of three independent music therapy sessions. Each point refers to the score of the observed well-being, representing the adjacent 30 seconds. The blue line is the mean level of the baseline and music therapy phases. MT 1 = Music Therapy session one. MT 5 = Music Therapy session five. MT 10 = Music Therapy session 10. Here, the return to baseline between sessions is visually discernible.

well-being (see Supplementary figures. Exceptions were "Greg," Session 10 and "Iris," Session 10). This indicates more experimental control and an increase in explanatory power of these results. For sociable interactions, no such consistent pattern was present.

Meta-analysis

Robust cluster variance estimation meta-analysis with small sample correction showed that the well-being (OWLS) increased by 48% in the music therapy compared to regular social interaction (*LRR* =0.39, 95% CI = [0.28– 0.51], SE = 0.05, t (9.88) = 7.49, p = <0.001***, I^2 = 86.03, τ^2 = 0.04). Furthermore, as the different observations were nested within cases and not independent, sensitivity analyses were conducted to investigate the stability of the results. When testing for different rho's (correlations), the results were stable for different dependencies within the clusters.

Sociable interaction (VNVIS-CR) significantly increased with 32% in the music therapy (*LRR* = 0.28, 95% CI = [0.05– 0.50], SE = 0.10, t (9.97) = 2.78, p = <0.02*, I^2 = 90.62, τ^2 = 0.16). The subscale *verbal* sociable interaction decreased by −12% (*LRR* = −0.13, 95% CI = [−0.21– 0.05], SE = 0.03, t(8.85) = −3.86, p = <0.01**, I^2 = 55.89, τ^2 = 0.01). The subscale *nonverbal* sociable interaction increased by 51% (*LRR* =0.41, CI = [0.16– 0.67], SE = 0.12, t(9.99) = 3.59, p = <0.01**, I^2 = 92.43, τ^2 = 0.20). Sensitivity analyses also supported stability.

Heterogeneity and meta-regression

A high degree of heterogeneity was present in the meta-analysis of both measures. τ^2 describes the underlying variance between sessions, and τ is expressed in the same metric as the effect size (LRR). I^2 is a measure of the percentage of variability in the effect sizes across the sessions that is attributed to heterogeneity rather than sampling error (Fisher & Tipton, 2017) and values over 75% are interpreted as large (Higgins, Thompson, Deeks, & Altman, 2003).

To investigate the sources of heterogeneity, separate meta-regression analyses were conducted for the moderators "dementia severity" and "depression." Using dementia severity as a factor (CDR, "very mild," "mild," and "moderately severe"), we found a larger change in the nonverbal sociable interaction-subscale in the group with "moderately severe dementia" (80% change, *LRR* = 0.59, 95% CI = [0.05– 1.12], SE = 0.19, t(4) = 3.05 p <.04*) compared to "very mild dementia" (7% change, *LRR* = 0.07, 95% CI = [<0.00 – .07], SE = 0.00, t (8.26) = 445833617076805.88, p = <0.001***). No difference in comparison with "very mild dementia" was found for "mild dementia" (18% change, *LRR* =

Table 2. Effect-sizes (LRR) per session for well-being, sociable interaction, and subscales verbal and nonverbal interactions.

		OWLS	VNVIS-CR		
Case	Session no	Well-being	Sociable interaction (ratio)	Subscale – verbal (ratio)	Subscale – nonverbal (ratio)
Ann	1	^{i}0.24 (SE = 0.06) 27%	0.10 (SE = 0.17), 11%	−0.14 (SE = 0.08), −13%	0.19 (SE = 0.17), 21%
	5	0.10 (SE = 0.02) 10%	−0.03 (SE = 0.10), −3%	−0.23 (SE = 0.07), −21%	0.12 (SE = 0.11), 12%
	10	0.21 (SE = 0.03) 23%	−0.19 (SE = 0.04), – 17%	−0.17 (SE = 0.06), −15%	−0.15 (SE = 0.03), −14%
Beth	1	0.53 (SE = 0.13) 69%	^{i}0.58 (SE = 0.20), 79%	0.08 (SE = 0.19), 9%	^{i}0.62 (SE = 0.15), 87%
	5	0.55 (SE = 0.15) 73%	0.05 (SE = 0.21), 5%	−0.39 (SE = 0.19), −33%	0.39 (SE = 0.20), 48%
	10	0.59 (SE = 0.15) 80%	0.69 (SE = 0.12) 99%	0.10 (SE = 0.11), 11%	0.65 (SE = 0.12), 92%
Clare	1	0.24 (SE = 0.11) 27%	−0.01 (SE = 0.10), −1%	−0.05 (SE = 0.07), −5%	0.07 (SE = 0.08), 7%
	5	0.41 (SE = 0.08) 50%	−0.01(SE = 0.08), −1%	−0.11 (SE = 0.05), −10%	0.06 (SE = 0.14), 6%
	10	0.45 (SE = 0.08) 56%	0.02 (SE = 0.09), 2%	−0.04 (SE = 0.06), −4%	0.08 (SE = 0.09), 8%
Dina	1	0.20 (SE = 0.09) 22%	0.17 (SE = 0.16), 18%	−0.13 (SE = 0.05), −12%	0.31 (SE = 0.15), 37%
	5	0.25 (SE = 0.10) 28%	−0.02 (SE = 0.12), −2%	NA	−0.02 (SE = 0.12), −2%
	10	0.01 (SE = 0.08) 1%	−0.11 (SE = 0.04), −11%	−0.09 (SE = 0.04), −9%	−0.08 (SE = 0.06), −7%
Eric	1	0.65 (SE = 0.09) 92%	0.57 (SE = 0.09), 77%	−0.47 (SE = 0.12), −37%	0.93 (SE = 0.08), 155%
	5	0.34 (SE = 0.07) 40%	−0.26 (SE = 0.06), −23%	−0.63 (SE = 0.10), −47%	0.05 (SE = 0.07), 5%
	10	0.47 (SE = 0.09) 60%	0.02 (SE = 0.11), 2%	−0.42 (SE = 0.15), −35%	0.20 (SE = 0.13), 23%
Fred	1	0.40 (SE = 0.05), 49%	−0.02(SE = 0.17), −2%	−0.34 (SE = 0.08), −29%	0.25 (SE = 0.17), 29%
	5	0.70 (SE = 0.09), 102%	0.90 (SE = 0.14), 146%	−0.01 (SE = 0.11), −1%	1.14 (SE = 0.16), 212%
	10	0.48 (SE = 0.08), 61%	0.85 (SE = 0.13), 134%	−0.03 (SE = 0.14), −2%	0.99 (SE = 0.13), 169%
Greg	1	0.58 (SE = 0.10) 78%	0.09 (SE = 0.17), 9%	−0.41 (SE = 0.16), −34%	0.41 (SE = 0.14), 51%
	5	0.49 (SE = 0.10) 63%	0.43 (SE = 0.17), 54%	0.01 (SE = 0.17), 1%	0.43 (SE = 0.15), 53%
	10	0.21 (SE = 0.13) 23%	0.36 (SE = 0.15), 44%	0.21 (SE = 0.13), 23%	0.19 (SE = 0.11), 21%
Helen	1	0.42 (SE = 0.05) 52%	^{i}0.59 (SE = 0.21), 80%	0.02 (SE = 0.12), 2%	^{i}0.68 (SE = 0.20), 98%
	5	0.58 (SE = 0.07) 79%	0.51 (SE = 0.16), 66%	−0.14 (SE = 0.11), −13%	0.63 (SE = 0.16), 88%
	10	0.50 (SE = 0.09) 64%	0.07 (SE = 0.16), 8%	−0.09 (SE = 0.12), −9%	0.22 (SE = 0.13), 25%
Iris	1	0.15 (SE = 0.13) 16%	0.27 (SE = 0.22), 31%	0.02 (SE = 0.15), 2%	0.30 (SE = 0.19), 35%
	5	0.24 (SE = 0.08) 27%	0.04 (SE = 0.07), 4%	−0.11 (SE = 0.05), −10%	0.10 (SE = 0.08), 11%
	10	0.30 (SE = 0.09) 35%	0.41 (SE = 0.20), 51%	0.18 (SE = 0.15), 19%	0.32 (SE = 0.16), 38%
John	1	0.22 (SE = 0.07) 25%	−0.29 (SE = 0.10), −25%	−0.16 (SE = 0.15), −15%	−0.24 (SE = 0.08), −22%
	5	0.31 (SE = 0.04) 37%	0.60 (SE = 0.17), 81%	−0.07 (SE = 0.09), −7%	0.76 (SE = 0.17), 113%
	10	0.27 (SE = 0.06) 31%	−0.21 (SE = 0.11), −19%	−0.28 (SE = 0.11), −24	0.17 (SE = 0.13), 19%
Kate	1	0.81 (SE = 0.07) 125%	^{i}1.06 (SE = 0.13), 188%	−0.08 (SE = 0.11), −7%	^{i}1.36 (SE = 0.12), 288%
	5	0.54 (SE = 0.09) 71%	1.00 (SE = 0.11), 171%	−0.17 (SE = 0.09), −16%	1.27 (SE = 0.12), 258%

Abbreviations and interpretation: LRR = Log Response Ratio calculated with expected increase in values. SE = Standard Error. LRR is converted to % of change, and study-specific benchmarks for small effect is >20-49% change, medium effect is 50–70% change and large effect is >70% change. Sessions where increasing baseline-trend is detected is marked with ^{i}

0.17, 95% CI = [−0.07– 0.41], SE = 0.09, t(4) = 1.99 *p* = .12). In total, the absolute increase in the "moderately severe" group was 93% (LRR = 0.65, adding the intercept (0.07) to the subgroup effect (0.59)). Heterogeneity for this analysis was large as well (I^2 = 91.21, τ^2 = 0.18). For the well-being measure, the results indicated that dementia severity could predict the effect of music therapy as well, but the reliability of the results (degrees of freedom <4) were too low to be trustworthy (Tipton & Pustejovsky, 2015).

Pre-scores of "depression" from the NPI-Q (Kaufer et al., 2000; dichotomized to 0 = "no or mild depression," 1 = "moderate or severe depression") did not predict the effects of music therapy. Conducting meta-regression with time as a factor does not support our hypothesis of accumulated increases in well-being and sociable interaction over time.

Relative frequencies of behavioral expressions of well-being and sociable interaction

Table S2 shows the relative frequencies of all the items from the observational scales in the different phases. The largest changes for well-being from baseline to intervention on group-level were found in the items "enjoyment" (45.9%), "happiness" (23.2%), "express identity" (36.5%), "relationship" (34.3%), and "positive feedback" (14.1%).

The largest changes in sociable-interaction items were found in the items "positive affect" (33%), "calm/relaxed" (20.9%), "appears aloof" (−19.4%), "vocalize negative affect" (−10.6%), and "responds to questions" (−30.9%).

The frequency of the remaining items of both scales changed <10%.

Table 3. Results of secondary measures.

Measure	N	Median pre	Median post	W	p
VAMS positive emotions	N_o = 206	46	63.5	2640.5	<.001***
VAMS negative emotion	N_o = 206	7	4	38,538	<.001***
QoL-AD	N = 11	22	26	15	.72
RSS	N = 11	24	27	21.5	.674
NPI-Q	N = 11	12	8	36	.014*

N_o refer to the total repeated number of *observations* included in the data-analysis. Due to the skewness of the emotions, parametric testing including time was not possible. VAMS-items ranges from 0– 100 where higher values indicate more intense emotions. QoL-AD Scores range 13– 52 where higher scores indicate more well-being. RSS range from 0– 60 where higher scores indicate higher caregiver burden. NPI-Q range from 0– 36 where higher scores indicate presence of more severe neuropsychiatric symptoms.

Secondary measures

Results from secondary measurements are reported in Table 3. In the music therapist-led sessions, positive self-reported emotions significantly increased from pre- to post-music therapy sessions. Negative self-reported emotions significantly decreased, but a floor effect makes the clinical relevance of the lowered negative emotions uncertain.

From the pre- to post intervention period, we did not find support for our hypothesized change in self-reported long-term well-being or caregiver burden. However, the hypothesized stable neuropsychiatric symptoms did significantly decrease from before to after the intervention period.

Discussion

This study investigated individually tailored music therapy compared to regular social interaction for home-dwelling people living with dementia. Within sessions, we found close to a moderate (48%) increase in observed well-being. A significant increase in the self-rated positive emotions from pre- to post-sessions in our secondary measures supports these observations. A small (32%) increase in sociable behavior toward significant others was found within sessions, where people with moderate dementia severity showed a large (93%) increase in nonverbal sociable interaction during music therapy. We did not see an increase in long-term well-being or lowered caregiver burden at the end of treatment. Nevertheless, a significant within-person decrease in neuropsychiatric symptoms from pre- to post-treatment suggests some positive long-term effects.

The increase in momentary well-being is in line with recent meta-analyses of music therapy (van der Steen et al., 2018; Zhang et al., 2017). Randomized controlled trials of musical interventions have shown inconsistent results, with some studies finding increased well-being (Cho, 2018; Hsu et al., 2015; Särkämö et al., 2014) and others not reaching statistical significance (Raglio et al., 2015; Ridder et al., 2013) or not finding an effect (Cooke et al., 2010). Studies using narrower timeframes with measures focusing on momentary changes (Cho, 2018; Hsu et al., 2015), as well as utilizing single-case designs and repeated fine-grained measures (Clair, 2002; Schall et al., 2015) seem to optimize the evaluation of well-being.

When exploring the behavioral content of OWLS in the different phases, we saw a large increase in expressions of enjoyment and happiness during music therapy. This is of clinical relevance, as several participants expressed “experiencing more happiness” as the most important personalized goal in therapy. Furthermore, we observed a large increase in positive expressions of identity, an objective identified as valuable by home-dwelling people living with dementia (Reilly et al., 2020; von Kutzleben et al., 2012). While dementia may lead to loss of role-functions and disrupt one’s self-concept, personalized music may serve to maintain and support a feeling of identity (Baird & Thompson, 2018; McDermott, Orrell, & Ridder, 2014). Familiar music triggering valued autobiographical memories may contribute to a more positive self-perception (Brancatisano et al., 2020). People living with dementia have expressed engaging in social participation as an important goal in interventions (Øksnebjerg et al., 2018). We found a large increase in joint interaction and turn-taking during music therapy (“relationship”), resembling the presence of mutual engagement found in the single-case study of Clair (2002). This underlines the potential for music therapy to facilitate social interaction and connection (Brancatisano et al., 2020).

In the current approach, we emphasize *sociable* behavioral expressions. Meta-analyses have identified a small effect of music therapy in people living with dementia on behavioral disturbances, with measures solely focusing on non-sociable behaviors such as agitation (Abraha et al., 2017; van der Steen et al., 2018). When investigating the relative frequencies of the items in VNVIS-CR, we found increases mainly in the sociable as opposed to the non-sociable behaviors. This suggests assessing sociable behavioral expressions is important for detecting behavioral effects from music therapy (Dowson et al., 2019; van der Steen et al., 2018).

Experiencing connections is identified as important for people living with dementia (McDermott et al., 2014; Reilly et al., 2020). At the group level, we found a large decrease in apathy ("appears aloof"). These changes illustrate how using individually tailored music may enable social capacities in people living with dementia (Brancatisano et al., 2020), potentially enhancing dyadic relationships. This is clinically important, as lack of mutuality is associated with caregiver burden (Cheng, 2017). While caregiver burden was stable from pre- to post-intervention, inspecting the logs in the study showed overall positive experiences of collateral-led sessions.

The slight inverse effect (−12%) of verbal sociable interaction may be explained by the low verbal activity in the music intervention compared to the higher verbal activity in the baseline phase, as the difference mainly concerned responding to questions and using coherent and relevant verbal communication.

Heterogeneity

The high degree of unexplained heterogeneity indicates that the effect forwell-being and sociable interactions varied to a large extent from session to session, both within and across the participants. Except for dementia severity, no other significant predictors were found. However, with increasing dementia severity, the potential for change may be greater as the social impairment is more prominent (Livingston et al., 2017). It seems evident that as dementia progresses, more focus on nonverbal positive expressions are important.

Symptoms of BPSD are known to fluctuate in people living with dementia (Kales et al., 2015) and may have indirectly contributed to the varying levels of well-being and sociable interaction.

Strengths and limitations

This single-case design intervention was conducted in a natural setting with optimal ecological validity and a meaningful comparison through using the subjects as their own controls. Offering individualized and preference-based music as opposed to pre-selected musical alternatives should also be considered a strength of this study. The lack of blinding introduces a risk of bias but is difficult to apply in single-case designs (Manolov et al., 2014). Meanwhile, the assessors were independent of the music therapists. Additionally, the intra-subject replication increases internal validity, and the inter-subject replication and meta-analytic approach together increase the external validity (Manolov et al., 2014). Replications in other research sites with other researchers will increase generalizability (Kratochwill et al., 2012). The observational measures (Williams et al., 2017; Madsø et al., manuscript under review) need further validation in different studies and contexts.

Although causal inferences are only possible in strictly controlled experimental designs, one may still imply that the intervention is the most plausible explanation of the changes from baseline to intervention in a single-case design (Kratochwill et al., 2012). A systematic return to baseline in the observations of well-being between sessions increases the likelihood of experimental control of these results. Other competing explanations for the change from baseline to intervention are monotonic trends, maturation, or history (Ledford & Gast, 2009; Tarlow, 2016). Future single-case studies should use multiple-baseline or equivalent randomized designs for investigating sociable interaction (Manolov et al., 2014). Continuing the observations while adding a post-intervention phase during the same day (ABA) could provide an opportunity to investigate whether the increase in expressions of well-being and sociable interaction are short-lived. However, continuing the video observation could also exhaust the participants and terminating the observation in a baseline-phase challenge ethical standards (Ledford & Gast, 2009).

The baseline-phase was only of 5 minutes duration, and we recommend that future studies should increase the baseline length to at least 15– 20 minutes in order to secure proper baseline stability (Ledford & Gast, 2009). Using a pre-defined length of all baseline-phases, instead of changing phase when stability or trend was established, is a weakness of our design (Ledford & Gast, 2009). The detection of a baseline trend would be a possible threat to internal validity, as stability in the baseline phase is important to validly compare phases (Manolov et al., 2014). Still, there was only an increasing baseline trend in one session when using OWLS and three sessions using VNVIS-CR, and thus, the baseline trend as such seems to be of little concern for the interpretation of the overall results in this meta-analysis.

The variability of the scores of the OWLS and VNVIS-CR within each session resembles the results of Schall et al. (2015) and suggests that the constructs we measured are naturally fluctuating in this population.

Even though we planned for the dyads to use music between sessions, their logs showed varying completion of this part of the intervention. A more structured approach toward the implementation of the caregiver-led music session may have helped, as conducted by Clair (2002) and Baker et al. (2012).

Conclusion

Individually tailored music therapy did positively impact short-term well-being in the care recipients. In addition, the potential of music therapy to increase sociable behavior toward caregivers warrants further investigation. The behaviors observed in the single-case design are evaluated as relevant for people living with dementia (Øksnebjerg et al., 2018; Reilly et al., 2020), following the advice to have a primary focus on the positive effects music therapy may offer (Dowson et al., 2019; van der Steen et al., 2018). The varying effect of music therapy from session to session calls for future studies to investigate sources of heterogeneity more closely. Here, single-case designs with high ecological validity may be a feasible approach. Observing and measuring covariates may hopefully increase the precision of the prediction of effects from music therapy for well-being and sociable interaction. Such covariates may include coding musical elements of the intervention, relevant symptoms, or elements in the communication from the caregiver or therapist.

Notes

1. Two primary measures were changed after pilot testing, and deviates from the pre-registered protocol. CODEM-instrument (Kuemmel, Haberstroh, & Pantel, 2014) was changed to Verbal and Nonverbal Interaction Scale – Care Recipient (VNVIS- CR, Williams et al., 2017). CODEM measures communication behavior, and the underlying construct resembles VNVIS-CR. Observed Emotion Rating Scale (OERS, Lawton, Van Haitsma, Perkinson, & Ruckdeschel, 1999) was replaced with Observable Well-being in Living with Dementia-Scale (OWLS, manuscript in preparation) the latter focusing solely on positive behavioral expressions
2. Contact corresponding author for details about the Observable Well-being in Living with Dementia-Scale (OWLS)

Clinical implications

- In-home music therapy has the potential to increase momentary well-being and sociable interactions for people living with dementia
- Given the relational impact of dementia, including family caregivers in music therapy interventions may support the dyadic need for relation reciprocity

Acknowledgments

We are grateful to all participants for sharing their experiences with us during this project, and the music therapists' valuable work and engagement. Also, we would like to thank Professor Nancy Pachana for helpful advice in the final stage of the writing. Professor Maurice Mittelmark and psychology student Sofie Holt Korneliussen was involved with translation of

scales, with the latter also involved in coding of videos together with psychology student Marte Muri.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Meltzers Research Foundation [811425]; POLYFON Knowledge Cluster of Music Therapy [811305]; NKS Olaviken Gerontopsychiatric Hospital; University of Bergen.

ORCID

Kristine Gustavsen Madsø CandPsychol http://orcid.org/0000-0002-4371-7576

References

Abraha, I., Rimland, J. M., Trotta, F. M., Dell'Aquila, G., Cruz-Jentoft, A., Petrovic, M., … Cherubini, A. (2017). Systematic review of systematic reviews of non-pharmacological interventions to treat behavioural disturbances in older patients with dementia. The SENATOR-OnTop series. *BMJ Open*, *7*(3), e012759–e012759. doi:10.1136/bmjopen-2016-012759

Algar, K., Woods, R. T., & Windle, G. (2016). Measuring the quality of life and well-being of people with dementia: A review of observational measures. *Dementia*, *15*(4), 832–857. doi:10.1177/1471301214540163

Baird, A., & Thompson, W. F. (2018). The impact of music on the self in dementia. *Journal of Alzheimer's Disease*, *61*(3), 827–841. doi:10.3233/JAD-170737

Bakeman, R., & Quera, V. (2012). Behavioral observation. In H. Cooper, M. Paul, L. L. Debra, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology* (pp. 207–225). New York: American Psychological Association.

Baker, F. A., Grocke, D., & Pachana, N. A. (2012). Connecting through music: A study of a spousal caregiver-directed music intervention designed to prolong fulfilling relationships in couples where one person has dementia. *The Australian Journal of Music Therapy*, *23* (2012), 4–21.

Brancatisano, O., Baird, A., & Thompson, W. F. (2020). Why is music therapeutic for neurological disorders? The therapeutic music capacities model. *Neuroscience and Biobehavioral Reviews*, *112*, 600–615. doi:10.1016/j.neubiorev.2020.02.008

Camic, P. M., Williams, C. M., & Meeten, F. (2013). Does a 'singing together group' improve the quality of life of people with a dementia and their carers? A pilot evaluation study. *Dementia*, *12*(2), 157–176. doi:10.1177/1471301211422761

Cheng, S.-T. (2017). Dementia caregiver burden: A research update and critical analysis. *Current Psychiatry Reports*, *19* (9), 1–8. doi:10.1007/s11920-017-0818-2

Cho, H. K. (2018). The effects of music therapy-singing group on quality of life and affect of persons with dementia: A randomized controlled trial. *Frontiers in Medicine*, *5*. doi:10.3389/fmed.2018.00279

Clair, A. A. (2002). The effects of music therapy on engagement in family caregiver and care receiver couples with dementia. *American Journal of Alzheimer's Disease and Other Dementias*, *17*(5), 286–290. doi:10.1177/153331750201700505

Clark, I. N., Tamplin, J. D., & Baker, F. A. (2018). Community-dwelling people living with dementia and their family caregivers experience enhanced relationships and feelings of well-being following therapeutic group singing: A qualitative thematic analysis. *Frontiers in Psychology*, *9*, 1332. doi:10.3389/fpsyg.2018.01332

Cooke, M., Moyle, W., Shum, D., Harrison, S., & Murfield, J. (2010). A randomized controlled trial exploring the effect of music on quality of life and depression in older people with dementia. *Journal of Health Psychol*, *15*(5), 765–776. doi:10.1177/1359105310368188

Davidson, J. W., & Fedele, J. (2011). Investigating group singing activity with people with dementia and their caregivers: Problems and positive prospects. *Musicae Scientiae*, *15*(3), 402–422. doi:10.1177/1029864911410954

Dewing, J. (2007). Participatory research: A method for process content with persons who have dementia. *Dementia*, *6* (1), 11–25. doi:10.1177/1471301207075625

Dowson, B., McDermott, O., & Schneider, J. (2019). What indicators have been used to evaluate the impact of music on the health and wellbeing of people with dementia? A review using meta-narrative methods. *Maturitas*, *127*(2019), 26–34. doi:10.1016/j.maturitas.2019.06.001

Dröes, R. M., Chattat, R., Diaz, A., Gove, D., Graff, M., Murphy, K., … the INTERDEM sOcial Health Taskforce. (2017). Social health and dementia: A European consensus on the operationalization of the concept and directions for research and practice. *Aging & Mental Health*, *21*(1), 4–17.

Fisher, Z., & Tipton, E. (2017). *robumeta: Robust variance meta-regression*. Retrieved from https://CRAN.R-project.org/package=robumeta

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. doi:10.1016/0022-3956(75)90026-6

Fusar-Poli, L., Bieleninik, Ł., Brondino, N., Chen, X.-J., & Gold, C. (2018). The effect of music therapy on cognitive functions in patients with dementia: A systematic review and meta-analysis. *Aging & Mental Health*, *22*(9), 1103–1112. doi:10.1080/13607863.2017.1348474

Greene, J. G., Gardiner, M., & Timbury, G. C. (1982). Measuring behavioral disturbance of elderly demented patiens in the community and its effects on relatives: a factor analytic study. *Age and ageing*, *11*, 121–126. 10.1093/ageing/11.2.121

Hedges, L. V. (2019). Stochastically dependent effect sizes. In C. Harris, L. V. Hedges, & J. C. Valentine (Eds.), *Handbook of research synthesis and meta-analysis* (3rd ed., pp. 281–298). New York: Russel Sage Foundation.

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. doi:10.1136/bmj.327.7414.557

Hsu, M. H., Flowerdew, R., Parker, M., Fachner, J., & Odell-Miller, H. (2015). Individual music therapy for managing neuropsychiatric symptoms for people with dementia and their carers: A cluster randomised controlled feasibility study. *BMC Geriatrics*, *15*(1), 84. doi:10.1186/s12877-015-0082-4

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, *240*, 565–572.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews*, *10*(3), 235–266. doi:10.1016/j.plrev.2013.05.008

Kales, H. C., Gitlin, L. N., & Lyketsos, C. G. (2015). Assessment and management of behavioral and psychological symptoms of dementia. *BMJ*, *350*(mar02 7), 1–16. doi:10.1136/bmj.h369

Kaufer, D. I., Cummings, J. L., Ketchel, P., Smith, V., MacMillan, A., Shelley, T., … DeKosky, S. T. (2000). Validation of the NPI-Q, a brief clinical form of the neuropsychiatric inventory. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *12*(2), 233–239. doi:10.1176/jnp.12.2.233

Kitwood, T. (1997). *Dementia reconsidered: The person comes first*. Buckingham: Open University Press.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2012). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26–38. doi:10.1177/0741932512452794

Kuemmel, A., Haberstroh, J., & Pantel, J. (2014). CODEM instrument: Developing a tool to assess communication behavior in dementia. *The Journal of Gerontopsychology and Geriatric Psychiatry*, *27*(1), 23–31. doi:10.1024/1662-9647/a000100

Lawton, M. P., Van Haitsma, 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*(1), 69–81.

Ledford, J. R., & Gast, D. L. (Eds.). (2009). *Single subject research methodology in behavioral sciences*. Londin: Taylor & Francis Group.

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., … Mukadam, N. (2017). Dementia prevention, intervention, and care. *Lancet*, *390*(10113), 2673–2734. doi:10.1016/S0140-6736(17)31363-6

Logsdon, R. G., Gibbons, L. E., & Teri, L. (1999). Quality of Life in Alzheimer's disease. *Journal of Mental Health and Aging*, *5*, 21–32. 10.1159/000090744

Madsø, K. G., Pachana, N. A., & Nordhus, I. H. (manuscript under review). *Development of the Observable Well-being in Living with Dementia-Scale (OWLS): a measure for well-being during psychosocial and creative interventions* Department of Clinical Psychology, University of Bergen.

Manolov, R., Gast, D. L., Perdices, M., & Evans, J. J. (2014). Single-case experimental designs: Reflections on conduct and analysis. *Neuropsychological Rehabilitation:An International Journal*, *24*(3–4), 634–660. doi:10.1080/09602011.2014.903199

McDermott, O., Orrell, M., & Ridder, H. M. (2014). The importance of music for people with dementia: The perspectives of people with dementia, family carers, staff and music therapists. *Aging & Mental Health*, *18*(6), 706–716. doi:10.1080/13607863.2013.875124

Miranda-Castillo, C., Woods, B., & Orrell, M. (2013). The needs of people with dementia living at home from user, caregiver and professional perspectives: A cross-sectional survey. *BMC Health Service Research*, *13*(1), 43. doi:10.1186/1472-6963-13-43

National Institute for Healt and Care Excellence. (2018). *Dementia: Assessment, management and support for people living with dementia and their carers [NG97]*. Retrieved from https://www.nice.org.uk/guidance/ng97/

Noldus Information Technology. (2015). *Noldus observer XT, 12.5*. Author. Retrieved from https://www.noldus.com/

Øksnebjerg, L., Diaz-Ponce, A., Gove, D., Moniz-Cook, E., Mountain, G., Chattat, R., & Woods, B. (2018). Towards capturing meaningful outcomes for people with dementia in psychosocial intervention research: A pan-European consultation. *Health Expectations*, *21*(6), 1056–1065. doi:10.1111/hex.12799

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, *35*(4), 303–322. doi:10.1177/0145445511399147

Pustejovsky, J. E. (2015). Measurement-comparable effect sizes for single-case studies of free-operant behavior. *Psychological Methods*, *20*(3), 342–359. doi:10.1037/met0000019

Pustejovsky, J. E. (2018). Using response ratios for meta-analyzing single-case designs with behavioral outcomes. *Journal of School Psychology*, *68*, 99–112. doi:10.1016/j.jsp.2018.02.003

Pustejovsky, J. E. (2019). Procedural sensitivities of effect sizes for single-case designs with directly observed behavioral outcome measures. *Psychological Methods*, *24*(2), 217–235. doi:10.1037/met0000179

Pustejovsky, J. E., & Swan, D. M. (2018). *Single-case effect size calculator (Version 0.5) Web application*. Retrieved from https://jepusto.shinyapps.io/SCD-effect-sizes/

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raglio, A., Bellandi, D., Baiardi, P., Gianotti, M., Ubezio, M. C., Zanacchi, E., … Stramba-Badiale, M. (2015). Effect of active music therapy and individualized listening to music on dementia: A multicenter randomized controlled trial. *Journal of American Geriatrics Society*, *63*(8), 1534–1539. doi:10.1111/jgs.13558

Reilly, S. T., Harding, A. J. E., Morbey, H., Ahmed, F., Williamson, P. R., Swarbrick, C., … Keady, J. (2020). What is important to people with dementia living at home? A set of core outcome items for use in the evaluation of non-pharmacological community-based health and social care interventions. *Age and Ageing*, *49*(4), 664–671. doi:10.1093/ageing/afaa015

Ridder, H. M. O., Stige, B., Qvale, L. G., & Gold, C. (2013). Individual music therapy for agitation in dementia: An exploratory randomized controlled trial. *Aging & Mental Health*, *17*(6), 667–678. doi:10.1080/13607863.2013.790926

Rolvsjord, R., Gold, C., & Stige, B. (2005). Research rigour and therapeutic flexibility: Rationale for a therapy manual developed for a randomised controlled trial. *Nordic Journal of Music Therapy*, *14*(1), 15–32. doi:10.1080/08098130509478122

RStudio Team. (2020). *RStudio: Integrated development for R*. PBC. Retrieved from http://www.rstudio.com/ .

Särkämö, T. (2018). Music for the ageing brain: Cognitive, emotional, social, and neural benefits of musical leisure activities in stroke and dementia. *Dementia*, *17*(6), 670–685. doi:10.1177/1471301217729237

Särkämö, T., Tervaniemi, M., Laitinen, S., Numminen, A., Kurki, M., Johnson, J. K., & Rantanen, P. (2014). Cognitive, emotional, and social benefits of regular musical activities in early dementia: Randomized controlled study. *The Gerontologist*, *54*(4), 634–650. doi:10.1093/geront/gnt100

Schall, A., Haberstroh, J., & Pantel, J. (2015). Time series analysis of individual music therapy in dementia. Effects on communication behavior and emotional well-being. *The Journal of Gerontopsychology and Geriatric Psychiatry*, *28*(3), 113–122. doi:10.1024/1662-9647/a000123

Steingrimsdottir, H. S., & Arntzen, E. (2015). On the utility of within-participant research design when working with patients with neurocognitive disorders. *Clinical Interventions in Aging*, *10*, 1189–1199. doi:10.2147/CIA.S81868

Stern, R. A., Arruda, J. E., Hooper, C. R., Wolfner, G. D., & Morey, C. E. (1997). Visual analogue mood scales to measure internal mood state in neurologically impaired patients: Description and initial validity evidence. *Aphasiology*, *11*(1), 59–71. doi:10.1080/02687039708248455

Strobel, C., & Engedal, K. (2008). *MMSE-NR. Norsk Revidert Mini Mental Status Evaluering. Revidert og utvidet manual.* Oslo: Nasjonalt kompetansesenter for aldring og helse.

Tarlow, K. R. (2016). An improved rank correlation effect size statistic for single-case designs: Baseline corrected tau. *Behavior Modification*, *41*(4), 427–467. doi:10.1177/0145445516676750

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., … Wilson, B. (2016). The single-case reporting guideline in behavioural interventions (SCRIBE) 2016 statement. *Canadian Journal of Occupational Therapy (1939)*, *83*(3), 184–195. doi:10.1177/0008417416648124

Terwee, C. B., Bot, S. D. M., De Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., … De Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. doi:10.1016/j.jclinepi.2006.03.012

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634. doi:10.3102/1076998615606099

van der Steen, J. T., Smaling, H. J., van der Wouden, J. C., Bruinsma, M. S., Scholten, R. J., & Vink, A. C. (2018). Music-based therapeutic interventions for people with dementia. *Cochrane Database of Systematic Reviews*, *2018*(7). doi:10.1002/14651858.CD003477.pub4

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling and Development*, *93*(4), 403–411. doi:10.1002/jcad.12038

von Kutzleben, M., Schmid, W., Halek, M., Holle, B., & Bartholomeyczik, S. (2012). Community-dwelling persons with dementia: What do they need? What do they demand? What do they do? A systematic review on the subjective experiences of persons with dementia. *Aging & Mental Health*, *16*(3), 378–390. doi:10.1080/13607863.2011.614594

Wickham, H. (2016). *ggolot2: Elegant graphics for data analysis*. Springer-Verlag. Retrieved from https://ggplot2.tidyverse.org

Williams, C. L., Newman, D., & Hammar, L. M. (2017). Preliminary psychometric properties of the verbal and nonverbal interaction scale: An observational measure for communication in persons with dementia. *Issues in Mental Health Nursing*, *38*(5), 381–390. doi:10.1080/01612840.2017.1279248

World Health Organization. (2017). *Global action plan on the public health response to dementia 2017–2025*. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/259615/9789241513487-eng.pdf;jsessionid=961E40CEC20469345AD209F73642F2ED?sequence=1

Zhang, Y., Cai, J., An, L., Hui, F., Ren, T., Ma, H., & Zhao, Q. (2017). Does music therapy enhance behavioral and cognitive function in elderly dementia patients? A systematic review and meta-analysis. *Ageing Research Reviews*, *35* (2017), 1–11. doi:10.1016/j.arr.2016.12.003

Graphic design: Communication Division, UIB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230855744 (print)
9788230843505 (PDF)