ACTA MÉDICA PORTUGUESA
A Revista Científica da Ordem dos Médicos

ARTIGO ORIGINAL

A Preliminary Psychometric Case Study of the Music in Dementia Assessment Scales: European Portuguese Version (Midas-PT)

Um Estudo de Caso Psicométrico Preliminar da Music in Dementia Assessment Scales: Versão em Português Europeu (Midas-PT)

Lídia SOUSA[1,2,3], Bárbara MOURA[3], Orii MCDERMOTT[4], Lia FERNANDES[1,2,5]
Acta Med Port (In Press) ▪ https://doi.org/10.20344/amp.18763

ABSTRACT

Introduction: Music in Dementia Assessment Scales (MiDAS) is a standardized outcome measure aiming to capture the effects of music-based interventions in patients with dementia. It is a unique instrument regarding its specificity, with the potential to enhance research in the field of music in dementia care. The aim of this study was to report a preliminary psychometric study of the translated and adapted instrument to European Portuguese (MiDAS-PT).

Material and Methods: Care home residents with dementia attended bi-weekly group music-based interventions, for five weeks. Intervention coordinators and care home staff completed MiDAS ratings at every session and the Quality-of-Life Scale (QoL-AD) at three time-points. Inter-rater reliability, test-retest reliability, internal consistency, concurrent validity (QoL-AD) and construct validity were evaluated.

Results: A total of 529 forms were completed (staff = 235, therapist = 294). Low therapist inter-rater and test-retest reliability, good internal consistency, low concurrent validity, and good construct validity were found. There were high factor loadings between the five MiDAS items (Interest, Response, Initiation, Involvement, and Enjoyment).

Conclusion: This preliminary investigation indicated acceptable psychometric properties on a range of attributes, but more research is needed in order to definitely establish the psychometric value of the scale.

Keywords: Dementia/therapy; Music Therapy; Neuropsychological Tests; Portugal; Psychometrics; Reproducibility of Results; Surveys and Questionnaires

RESUMO

Introdução: A *Music in Dementia Assessment Scales* (MiDAS) é um instrumento estandardizado para avaliar os efeitos de intervenções terapêuticas baseadas na música em doentes com demência. Dada a sua especificidade e características, é uma escala particularmente interessante e importante na investigação do uso terapêutico da música na demência. O objetivo deste estudo foi reportar um estudo psicométrico preliminar do instrumento traduzido e adaptado para Português Europeu (MiDAS-PT).

Material e Métodos: Um grupo de doentes com demência, residentes em lar, frequentou sessões de intervenções terapêuticas baseadas na música em grupo, bissemanais, durante cinco semanas. Os coordenadores da intervenção e profissionais de saúde do lar aplicaram a MiDAS em todas as sessões e a *Quality-of-Life Scale* (QoL-AD) em três momentos. Foram avaliadas a fiabilidade inter-observadores, a fiabilidade teste-reteste e a validade de construto.

Resultados: Foram completados um total de 529 formulários (funcionários = 235, terapeutas = 294). Foram identificadas baixa fiabilidade inter-observadores e teste-reteste, boa consistência interna, baixa validade concorrente e boa validade de construto. Verificou-se uma elevada correlação entre os cinco itens da escala (Interesse, Resposta, Iniciativa, Envolvimento e Satisfação) na análise fatorial.

Conclusão: Esta investigação preliminar demonstrou propriedades psicométricas aceitáveis em variados atributos, sendo necessária investigação adicional para estabelecer definitivamente o valor psicométrico do instrumento.

Palavras-chave: Demência/tratamento; Inquéritos e Questionários; Musicoterapia; Portugal; Psicometria; Reprodutibilidade dos Testes; Testes Neuropsicológicos

INTRODUCTION

People living with dementia (PwD) experience not only cognitive symptoms but also less well known behavioral and psychological symptoms of dementia (BPSD), like agitation, disinhibition, irritability, and psychosis. The management of those symptoms remains an exceptional challenge for clinicians and carers.[1] The frequent use of sedatives and physical restraint is associated with sub-optimal efficacy and important adverse effects, such as increased risk of cardiovascular events and mortality.[2-9] Non-pharmacological interventions are recommended as first line approaches for BPSD prevention and management,[10-13] but its widespread applicability is not yet fully established.[14-20]

Music-based interventions (Mbi) – including formal music therapy and other less strictly defined therapeutic music activities, which are performed for the purpose of obtaining health benefits, but without the intervention of a trained music therapist – seem to have a positive effect in mood and behavior of PwD, at least in the short term.[15,18,21,22] Various forms of Mbi have become increasingly popular in dementia care – highlighting that music is widely accepted as beneficial for

1. Department of Clinical Neurosciences and Mental Health. Faculty of Medicine. Porto University. Porto. Portugal.
2. Center for Health Technologies and Services Research. Porto. Portugal.
3. Psychiatry Department. Centro Hospitalar de Vila Nova de Gaia-Espinho. Vila Nova de Gaia. Portugal.
4. Faculty of Medicine & Health Sciences. University of Nottingham. Nottingham. United Kingdom.
5. Psychiatry Department. Centro Hospitalar Universitário de S. João. Porto. Portugal.
✉ **Autor correspondente:** Lídia Sousa. lidiasousa99@gmail.com
Recebido/*Received*: 26/06/2022 - **Aceite/*Accepted*:** 07/10/2022 - **Publicado *Online*/*Published Online*:** 13/01/2023

the general wellbeing of PwD, potentially helping them to adapt, enabling the connection with people around them, the environment and society[17] – but robust evidence on their effectivity is lacking.

Mbi are especially difficult to describe fully and transparently, because of particular aspects such as the complexity of the music stimuli and the interpersonal component of the intervention.[23] The scarcity of standardized and adequate outcome measures also contributes to the dearth of robust evidence.[24] Author derived non-validated instruments are frequently used, and standardized measures are commonly derived from other disciplines. Those allow for comparisons between the effects of Mbi and other interventions but might fail in capturing the essence of the Mbi effects and what can specifically be created and/or changed for PwD.[19]

There is a need for robust, standardized, dementia specific instruments, that allow researchers to capture a holistic picture of the therapy process and effects of Mbi, specially tailored to be used with patients with major cognitive deficits. The Music in Dementia Assessment Scales (MiDAS) has been specifically developed for this purpose.[25] Additionally, it does not focus exclusively on capturing the reduction of clinical symptoms – like most non-music therapy outcome measures – but encompasses a wider range of well-being indicators (Interest, Response, Initiation, Involvement, and Enjoyment).

In previous work, we reported the translation and cross-cultural adaptation of MiDAS to European Portuguese.[26] Face and content validity were examined at this stage. The MiDAS Portuguese Version (MiDAS-PT) was approved by the authors of the original instrument and is publicly available at http://www.midas.aau.dk.

We aimed to report a preliminary psychometric study of MiDAS-PT, in order to validate its use for research and clinical purposes.

MATERIAL AND METHODS

This preliminary psychometric study is an exploratory work from a larger PhD project aiming to investigate the administration of Mbi to PwD in the acute setting.

MiDAS was administered repeatedly to PwD who participated in a Mbi, in order to investigate inter-rater reliability, test-retest reliability, internal consistency and construct validity. Concurrent validity was assessed using the Quality of Life in Alzheimer's disease scale (QoL-AD),[27] akin to what has been done for the original MIDAS version psychometric study.[28]

Two psychiatry trainees with experience in dementia care were responsible for delivering the Mbi and rating the MiDAS Therapist Version. Four registered nurses, who provided regular daily care to the residents, completed the MiDAS Staff ratings and Qol-AD.

Training and clarifications on how to use the instruments were delivered through a single individual meeting with a researcher.

Design

Bi-weekly group Mbi was provided to care home residents with dementia. Before a participant attended the Mbi, a staff member who was familiar with all the residents was asked to complete a MiDAS BEFORE rating based on the average presentation of the resident on the day. An "AFTER" form was completed by the same staff member a few hours after the Mbi.

Both intervention coordinators completed a "BEFORE" form and a "DURING" form immediately after each session. The "BEFORE" rating was based on the observation of the resident during the first five minutes of the session, while the "DURING" rating was based on the observation of the most clinically significant five minutes of that session.

The intervention coordinators and care home staff completed MiDAS ratings at every session. Quality of Life in Alzheimer's disease (QoL-AD)[27] was completed at baseline, mid-treatment, and end-treatment by a staff member.

Music-based intervention

The intervention consisted of semi-individualized sessions of music listening, lasting 30 to 45 minutes, during which patients were encouraged to sing, dance and engage with each other freely. Music was delivered through loudspeakers in a quiet room where patients had the option to sit in a circle. Musical preferences were assessed through informal interviews with the participants and their main carers. The most frequently chosen music types were: popular Portuguese music, traditional Portuguese music/folklore, fado and religious music. In total, 10 sessions were held, over a period of five weeks, and took place every Monday and Friday midmorning.

Participants

Elderly residents from one semi-private Portuguese care home were assessed for eligibility (n = 45). Being a typical Portuguese care home, located in an urban area, the institution offers non-specialized care to elders with and without

dementia. The inclusion criteria were having a clinical diagnosis of dementia and being capable of giving informed consent to participate (or having a legal representative). Significant hearing impairment was the only exclusion criteria applied. The recruitment process is illustrated in Fig. 1. From the 17 potential participants approached, 12 accepted to participate. Most were females (n = 10), widowed (n = 6) or married (n = 4), with a mean age of 79 years old (min - 69; max - 87). All had moderate to severe Alzheimer's type dementia and had been institutionalized for more than one year (mean admission time: 2 years; min: 1,5 years; max: 4 years).

Instruments

Music in Dementia Assessment Scales (MiDAS)

MiDAS consists of five visual analogue scales (VAS), each capturing a different dimension of the music intervention's effect: Interest, Response, Initiation, Involvement and Enjoyment.

The original English version was rigorously developed using qualitative methods and consensus approaches[25] and a preliminary psychometric evaluation indicates adequate psychometric properties on a range of attributes (high therapist inter-rater reliability; adequate staff test-retest reliability; adequate concurrent validity; and good construct validity) even though sample size was small.[28]

The response to each dimension of the scale is marked vertically on a 10 cm line and then converted to a score ranging from 0 to 100, with higher scores representing better effects. This may be illustrated with additional information regarding whether the person had some important reactions beyond the usual, through free text, and six optional behaviors can also be selected (agitated/aggressive; withdrawn/low in mood; restless/anxious; relaxed; attentive/interested; cheerful/smiling).

MiDAS has two versions, one to be completed by a staff member and the other for the music therapist. In the staff version, an assessment is carried out several hours "BEFORE" and "AFTER" the music therapy, on the same day, ideally by the same caregiver, who knows the person with dementia well. In the music therapist version, the music therapist marks the participant's response to the music therapy itself, with a retrospective assessment of the participant's behavior in both the beginning of the Mbi and at the clinically most relevant time "DURING" the intervention. The staff version can be used to assess the effect of music therapy based on the staff's subjective assessment and the music therapist version can provide more information about changes in the clinical context itself. The two forms should then be analyzed separately to mutually inform each other.[25]

Quality of life in Alzheimer's disease (QoL-AD)

QoL-AD[27] is a Likert scale, originally developed as a disease-specific scale for Alzheimer's disease, consisting of 13 items, including physical health, mood, family, ability to do chores around the house, and life as a whole. It can be used as a self-rating scale, or as a proxy measure completed by a carer, to evaluate a broad concept of well-being.

Statistical analysis

Statistical analysis was conducted with SPSS version 26 and R version 4.0.4. Inter-rater reliability, test-retest reliability, internal consistency, concurrent validity, and construct validity of MiDAS were evaluated, akin to the original MiDAS psychometric evaluation study.[28] Due to the small sample size, repeated observations of the same residents at different times were treated as unique observations when evaluating inter-rater reliability, concurrent validity, and construct validity. This approach was also used in the original MiDAS psychometric evaluation.

"Reliability analysis" consisted of the estimation of intraclass correlation coefficient (ICC) for inter-rater reliability and test-retest reliability. An ICC case 2 two-way random effects model with absolute agreement (McGraw and Wong, 1996)[29] was used. Since these analyses depend on a normal distribution of variables, normality tests were performed for Mbi first moment ("BEFORE") evaluations, excluding missing data. Asymmetry coefficient, Kurtosis coefficient and the Kolmogorov-Smirnov test were calculated.

Internal consistency was examined using Cronbach's α.

Concurrent validity was examined with Spearman's correlation coefficient.

Construct validity was evaluated through exploratory factor analysis. Both principal component analysis (PCA) and principal axis factoring (PAF), a descriptive method applicable when normality cannot be established, were performed. Two different approaches were used to estimate correlation coefficients. In factorial analysis Test 1, repeated observations of the same residents at different times were treated as unique observations and factorial analysis test 2 was conducted on the mean MiDAS scores of each participant. Due to the small sample size of this pilot study, squared multiple correlation (SMC) for each variable were used. To confirm the adequacy of the factorial analysis, visual inspection of the multiple scatter diagram, estimation of the diagonal values of the anti-image correlation matrix and calculation of Kaiser–Meyer–Olkin

(KMO) measure of sampling adequacy and Bartlett's test of sphericity[30] were performed. Squared multiple correlations (R^2) were also calculated, in order to determine the commonality and the adequacy of the factor analysis, according to the suggestion of Field.[31]

Ethical considerations

The present study complies with the Declaration of Helsinki and was approved by the ethics committee of the Faculty of Medicine of the University of Porto. All participants (or their legal representatives) provided written informed consent. Patients could decline to participate in any specific session, without the need for justification.

RESULTS

The 10 Mbi sessions were held as planned, but participant attendance varied, as illustrated in Table 1. As a substantial proportion of missing data (> 40%) was found, it was not possible to use 'multiple imputation' to work around the problem, as suggested by Jakobsen *et al*.[32]

A total of 529 MiDAS forms (staff = 235, therapist = 294) and 30 QoL-AD were completed. The mean of each of the five VAS items was calculated for both moments ("BEFORE" and "AFTER/DURING" respectively) as shown in Table 2. The mean total scores were: "BEFORE" 26.9 (SD 21.3); "AFTER/DURING" 60.3 (SD 30.9).

Reliability

The results of inter-rater reliability and test-retest reliability are shown on Table 3.

Since the asymmetry coefficient value was far from zero (0.27), as well as the Kurtosis coefficient (2.05), the distribution of the scores was not considered normal. The Kolmogorov-Smirnov test [D (575) = 0.09, $p < 0.001$], also confirmed this assumption. In order to normalize the symmetry and Kurtosis of the original scores, the square root transformation of the original data was conducted to evaluate ICC. Tests on original and transformed data were then performed to assess if there were any major differences between the results.

Inter-rater reliability: the analysis was carried out based on the observations produced by the two intervention coordinators. Since the differences between the results of the transformed data and the original data were not significant (ICC differences: range - 0.063 to 0.052, mean: 0.014, SD: 0.047) the original scores were retained. Analyzing the ICC scores ("BEFORE"), all items assumed values below 0.5, thus suggesting a low level of reliability. In the second moment of evaluation ("DURING"), the values were higher and statistically significant ($p < 0.05$), although in general still below the average level of reliability (0.6).

Test-retest reliability: 94 paired therapist forms were used, and all scores obtained by each pair for all contiguous sessions over time were compared. The differences between the results of the transformed and the original data were also not significant (ICC differences: range 0.07 to 0.14, mean: 0.1, SD: 0.027) and thus original scores were used. In this case, although the ICC assumes values higher than those obtained in the previous test, they are mostly below 0.6.

Internal consistency

Cronbach's α values revealed high correlations (α = 0.982) between the five VAS items (n = 457). The values regarding the initial observations of the intervention coordinators were 0.957 (n = 115), and for their second moment of evaluation 0.976 (n = 140). In the case of staff, the values were 0.974 (n = 117), "BEFORE" and 0.970 (n = 83), "AFTER".

Concurrent validity

QoL-AD was administered at the beginning (session 1), in the middle (session 5) and at the end (session 10) of the collection process, prior to the intervention. The three results (n = 30) were combined to assess the overall correlation, resulting in a Spearman ordinal correlation coefficient of 0.079 ($p = 0.714$). Similarly, Spearman's correlation coefficients for each of the three moments were neither acceptable nor significant: 0.566 (beginning), 0.116 (middle) and -0.008 (final), in relation to the scores assigned by intervention coordinators and staff ("BEFORE").

Construct validity

Factorial analysis test 1

This analysis was performed based on the scores attributed by the intervention coordinators ("BEFORE"), organized by item. Individual scores were treated as single observations. The visual inspection of the multiple scatter diagram of the five items allowed us to consider that the relationships between the variables with each other are approximately linear. The KMO value was 0.910, which exceeds the 0.6 value suggested by Pallant.[33] The Bartlett's sphericity test was statistically

significant ($p < 0.001$), thus allowing the decomposition into factors of the correlation matrix through a PCA. The diagonal values of the anti-image correlation matrix were all greater than 0.887, which is considered good, and therefore it was not necessary to remove any variables to perform the analysis.

PCA*:* the correlation matrix (Table 4) between the five items allowed us to verify that all correlations were statistically significant ($p < 0.01$) and greater than 0.773. The lowest correlation was between Initiation and Response. Factors (principal components) with eigenvalues greater than 1 were selected. Five components were extracted, but only the first was retained, which explained 85.6% of the variance. This fact allowed us to conclude about the existence of a single latent variable.

PAF*:* the results were similar to those obtained by PCA. The first factor explained 85.6% of the variance. The factor matrix had a very similar structure to the PCA component matrix: the item best associated with factor 1 was Interest and the worst associated was Initiation.

Factorial analysis test 2

This analysis was also performed based on the scores attributed by the intervention coordinators ("BEFORE"), organized by item. The averages of the assigned scores were calculated, per item. Due to the small sample size, R2-scores were calculated and since they were consistently higher than 0.5 for the five items of the VAS – varying between 0.924 and 0.980, with an average of 0.9616 – which exceeds the recommendation of the average level of communalities of 0.7 MacCallum *et al*,[34] a factorial analysis could be performed. The visual inspection of the multiple scatter diagram between the five items indicated the relationships between the variables were approximately linear. Regarding the quality of the correlations, the value of the KMO measure was 0.806, which exceeds the value of 0.6 suggested by Pallant.[33] The Bartlett's sphericity test was statistically significant ($p < 0.001$), thus allowing the decomposition into factors of the correlation matrix through the PCA. The diagonal values of the anti-image correlation matrix were all greater than 0.752, which is considered good, avoiding the need to remove any variable from the analysis.

The results of test 2 were similar to the previous factor analyses (Table 5). The first factor explained 95.8% of the variance and the component matrix had a very similar structure to that of the previous factor analyses: the item best associated with factor 1 was Interest and the worst associated was Initiation.

DISCUSSION

This preliminary psychometric evaluation of MiDAS-PT revealed good indicators regarding internal consistency and construct validity, with reliability and concurrent validity falling below expectations. However, some crucial aspects of the study methodology need to be cautiously considered, especially when comparing results with the psychometric study of the original version of MiDAS.

Inter-rater reliability and Test-retest reliability were analyzed through the calculation of ICC for the observations of the intervention coordinators only (since there weren't enough paired staff forms) – with a case 2 two-way random effects model with absolute agreement. The premise of absolute agreement might prevent the identification of agreement only at certain levels. The impossibility to analyze reliability for Staff – who know the PwD better and could thus be more consistent when rating MiDAS items – might have also impacted negatively on our reliability estimation.

The analysis of Table 3 indicates that reliability (inter-rater and test-retest) was always better in the second moment of observation, thus possibly indicating that the scale has better discriminative properties when there is some change in the basal state of the subjects – perhaps in the moment "BEFORE" the Mbi people with dementia were too withdrawn and apathetic to allow for fine discrimination in MiDAS items. For instance, regarding inter-rater reliability, almost all the scores in the second moment of observation were above the 0.5 threshold and the values reached statistical significance. Therefore, the scale might have better reliability for this second moment of evaluation.

Another important observation is that the test-retest reliability was significantly better than inter-rater reliability, for all items and in both moments of observations. This indication of better repeatability by the same observer than concordance between raters could be explained by the lack of training and proficiency of raters, since this is expected to affect the repeatability to less extent and the concordance between raters to greater extent. Importantly, in the second moment of observation, almost all items reached the value of 0.6, indicating a reasonable test-retest reliability.

In the preliminary psychometric study of the original version of MiDAS,[28] a low staff inter-rater reliability has also been reported. Despite this variability among raters, the test-retest reliability for this group was adequate, possibly indicating that the repeated use of the scale increased the consistency with which the instrument is used. A high inter-rater reliability among intervention coordinators was identified in this study, as expected, since they were quite familiar with the scale. We believe this corroborates the hypothesis that the scale is reliable, that appropriate training is provided.

Internal consistency measures the degree of correlations between the items on the same test. For this purpose, we calculated Cronbach's α values considering all observations and then separately, for therapist and staff and in both moments of observation. Results were similar for all the analyses and a very good internal consistency was admitted with all correlation coefficients being significantly above 0.8.

One observation that could help explain the counter-intuitive low inter-rater and test-retest reliability, with a good internal consistency according to Cronbach's α, is that the reliability tests used were based on the assumption of absolute agreement. Therefore, it is possible that there was some concordance that was not identified. There might be a correlation in the evaluations of different raters and in different ratings from the same rater, with better and worse scores being systematically attributed to the same patients, but without an absolute concordance in the scores. This is a point that deserves further exploration in future investigations, with appropriate methodology.

Concurrent validity investigates if a test correlates well with a measure that has previously been validated for the same construct, or, more often, presumably related constructs. In the psychometric study of the original version of MiDAS, QoL-AD has been used to evaluate this parameter. For replicability and comparison purposes we decided to use the same scale. In the original study, the concurrent validity of MiDAS in comparison with Qol-AD was only reasonable. This was not unexpected, since the scales were not designed to capture exactly the same construct. In our investigation of MiDAS-PT, concurrent validity was even lower. This might be related with the smaller size of our sample or with less training/familiarity of our raters with the scales. Notwithstanding, in our opinion, it would be preferable to use a more similar scale for concurrent validity, such as the recently developed Music Therapy Engagement scale for Dementia (MTED).[35] The latter evaluates musical experiences of PwD, while Qol-AD does in fact measure a slightly different construct and is not a scale prepared for repeated use in a short period of time.

Construct validity refers to the degree to which a test measures what it is supposed to be measuring – the construct that one wants to capture and measure. Factorial analysis is one of the methodologies commonly used to do so. Specifically, it enables the confirmation of the relationships between the test items and to identify the total number of dimensions represented on the test. It is commonly accepted that a good correlation between items and the identification of only one dimension indicates the scale has been properly developed in order to discriminate a certain construct. We performed two sets of factor analysis to accommodate the limitation of using repeated observations of the same residents as independent observations. As for the preliminary psychometric study of the original version of MiDAS, PCA and PAF were performed in original and transformed data and a second factorial test was performed using the average of the assigned scores per item of each participant, instead of treating all individual form scores as single observations. Different strategies confirmed the between item correlations and the one-dimensionality of the scale. Moreover, in order to be extra cautious, we hypothesized that the intervention could affect the interrelations between items, just like the mix of different raters (therapist and staff). In agreement with that hypothesis, we performed factorial analysis using only the forms from the "BEFORE" observation of intervention coordinators.

Finally, it can be argued that when the correlation between the different items of a scale is too high it can also indicate that they are too similar to discriminate their characteristics from each other. As explained by the original authors of MiDAS, having one construct (engagement with music) and using the total score of the five VAS items as the main outcome does not decrease the clinical relevance of MiDAS.

Strengths and limitations of the study

A rigorously translated and adapted Portuguese version of MiDAS was used, under the supervision of its original developers and of experts in dementia care. Statistical advice was obtained from a statistician with extensive experience in psychometric studies.

The main limitation of this study is its small sample size. Relatively low session attendance also decreased the power of the study. The fact that not all raters were previously familiar with the instruments can be a point of criticism as well, and one that helps contextualize the results and thus should be carefully addressed in future studies. Additionally, it was not possible to fully explore the reliability of staff rating.

The use of a more standardized therapeutic music intervention by a licensed music therapist could have produced different and more reproducible results, despite the fact that the MiDAS developers have suggested it could be used by non-music therapists

These are all points that should be better addressed in future studies in order to more accurately estimate the psychometric characteristics of the instrument. We suggest future studies to use evidence-based music therapy protocols to deliver Mbi as well as proficient raters of the scale. Subgroup analyses according to dementia severity (mild/moderated or severe) would also be interesting.

CONCLUSION

MiDAS-PT is the first instrument available in European Portuguese that was designed to evaluate the effects of Mbi in dementia care. This preliminary psychometric investigation suggests that MiDAS-PT has acceptable properties on a range of attributes even though the sample size was small and raters were not proficient users of the scale.

ACKNOWLEDGEMENTS

We would like to thank Sérgio Bacelar for proofreading the manuscript.

AUTHOR CONTRIBUTIONS

LS: Study concept and design, acquisition of data, data analysis and manuscript writing.
BM: Acquisition of data and preparation of the manuscript.
OM, LF: Analysis and interpretation of data and manuscript revision.

PROTECTION OF HUMANS AND ANIMALS

The authors declare that the procedures were followed according to the regulations established by the Clinical Research and Ethics Committee and to the Helsinki Declaration of the World Medical Association updated in 2013.

DATA CONFIDENTIALITY

The authors declare having followed the protocols in use at their working center regarding patients' data publication.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

FUNDING SOURCES

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

1. Gottesman RT, Stern Y. Behavioral and psychiatric symptoms of dementia and rate of decline in Alzheimer's disease. Front Pharmacol. 2019;10:1062.
2. Liperoti R, Pedone C, Corsonello A. Antipsychotics for the treatment of behavioral and psychological symptoms of dementia (BPSD). Curr Neuropharmacol. 2008;6:117-24.
3. Schneider LS, Dagerman KS, Insel P. Risk of death with atypical antipsychotic drug treatment for dementia: meta-analysis of randomized placebo-controlled trials. JAMA. 2005;294:1934-43.
4. Schneider LS, Tariot PN, Dagerman KS, Davis SM, Hsiao JK, Ismail MS, et al. Effectiveness of atypical antipsychotic drugs in patients with Alzheimer's disease. N Engl J Med. 2006;355:1525-38.
5. Ballard C, Waite J. The effectiveness of atypical antipsychotics for the treatment of aggression and psychosis in Alzheimer's disease. Cochrane Database Syst Rev. 2006;1:CD003476.
6. Gallini A, Andrieu S, Donohue JM, Oumouhou N, Lapeyre-Mestre M, Gardette V. Trends in use of antipsychotics in elderly patients with dementia: impact of national safety warnings. Eur Neuropsychopharmacol. 2014;24:95-104.
7. McConnell ES, Karel MJ. Improving management of behavioral and psychological symptoms of dementia in acute care: evidence and lessons learned from across the care spectrum. Nurs Adm Q. 2016;40:244-54.
8. Foebel AD, Onder G, Finne-Soveri H, Lukas A, Denkinger MD, Carfi A, et al. Physical restraint and antipsychotic medication use among nursing home residents with dementia. J Am Med Dir Assoc. 2016;17:184.e9-14.
9. Smeets CH, Gerritsen DL, Zuidema SU, Teerenstra S, van der Spek K, Smalbrugge M, et al. Psychotropic drug prescription for nursing home residents with dementia: prevalence and associations with non-resident-related factors. Aging Ment Health. 2018;22:1239-46.
10. International Psychogeriatric Association. The IPA Complete Guides to BPSD. 2012. [cited 2021 Dec 12]. Available from: https://www.ipa-online.org/publications/guides-to-bpsd.
11. Kales HC, Gitlin LN, Lyketsos CG. Assessment and management of behavioral and psychological symptoms of dementia. BMJ. 2015;350:h369.
12. National Institute for Health and Care Excelence. Dementia: supporting people with dementia and their carers in health and social care - Clinical Guideline. London: NICE; 2006.
13. Reus VI, Fochtmann LJ, Eyler AE, Hilty DM, Horvitz-Lennon M, Jibson MD, et al. The American Psychiatric Association practice guideline on the use of antipsychotics to treat agitation or psychosis in patients with dementia. Am J Psychiatry. 2016;173:543-6.
14. Abraha I, Rimland JM, Trotta FM, Dell'Aquila G, Cruz-Jentoft A, Petrovic M, et al. Systematic review of systematic reviews of non-pharmacological interventions to treat behavioural disturbances in older patients with dementia. The SENATOR-OnTop series. BMJ Open. 2017;7:e012759.
15. Chang YS, Chu H, Yang CY, Tsai JC, Chung MH, Liao YM, et al. The efficacy of music therapy for people with dementia: a meta-analysis of randomised controlled trials. J Clin Nurs. 2015;24:3425-40.
16. Livingston G, Kelly L, Lewis-Holmes E, Baio G, Morris S, Patel N, et al. Non-pharmacological interventions for agitation in dementia: systematic review of randomised controlled trials. Br J Psychiatry. 2014;205:436-42.
17. McDermott O, Charlesworth G, Hogervorst E, Stoner C, Moniz-Cook E, Spector A, et al. Psychosocial interventions for people with dementia: a synthesis of systematic reviews. Aging Ment Health. 2019;23:393-403.
18. McDermott O, Crellin N, Ridder HM, Orrell M. Music therapy in dementia: a narrative synthesis systematic review. Int J Geriatr Psychiatry. 2013;28:781-94.
19. van der Steen JT, Smaling HJ, van der Wouden JC, Bruinsma MS, Scholten RJ, Vink AC. Music-based therapeutic interventions for people with dementia.

Cochrane Database Syst Rev. 2018;7:CD003477.
20. Vink A, Hanser S. Music-based therapeutic interventions for people with dementia: a mini-review. Medicines. 2018;5:109.
21. Vasionyte I, Madison G. Musical intervention for patients with dementia: a meta-analysis. J Clin Nurs. 2013;22:1203-16.
22. Ueda T, Suzukamo Y, Sato M, Izumi S. Effects of music therapy on behavioral and psychological symptoms of dementia: a systematic review and meta-analysis. Ageing Res Rev. 2013;12:628-41.
23. Robb SL, Burns DS, Carpenter JS. Reporting guidelines for music-based interventions. Music Med. 2011;3:271-9.
24. Sousa L, Neves MJ, Moura B, Schneider J, Fernandes L. Music-based interventions for people living with dementia, targeting behavioral and psychological symptoms: a scoping review. Int J Geriatr Psychiatry. 2021;36:1664-90.
25. McDermott O, Orrell M, Ridder HM. The development of Music in Dementia Assessment Scales (MiDAS). Nord J Music Ther. 2015;24:232-51.
26. Sousa L, Moura B, McDermott O, Fernandes L. Portuguese translation and cultural adaptation of the Music in Dementia Assessment Scales (MiDAS). Nord J Music Ther. 2022;31:89-101.
27. Logsdon RG, Gibbons LE, McCurry SM, Teri L. Quality of life in Alzheimer's disease: patient and caregiver reports. J Ment Health Aging. 1999;5:21-32.
28. McDermott O, Orgeta V, Ridder HM, Orrell M. A preliminary psychometric evaluation of Music in Dementia Assessment Scales (MiDAS). Int Psychogeriatr. 2014;26:1011-9.
29. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1:30-46.
30. Bartlett MS. A note on the multiplying factors for various χ² approximations. J R Stat Soc Series B Stat Methodol. 1954;16:296-8.
31. Field A. Exploratory factor analysis. Discovering statistics using SPSS. 3rd ed. London: Sage; 2009.
32. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Med Res Methodol. 2017;17:162.
33. Pallant J. SPSS survival manual. Maidenhead: Open University Press; 2005.
34. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. Psychol Methods. 1999;4:84-99.
35. Tan J, Wee SL, Yeo PS, Choo J, Ritholz M, Yap P. A new music therapy engagement scale for persons with dementia. Int Psychogeriatr. 2019;31:49-58.

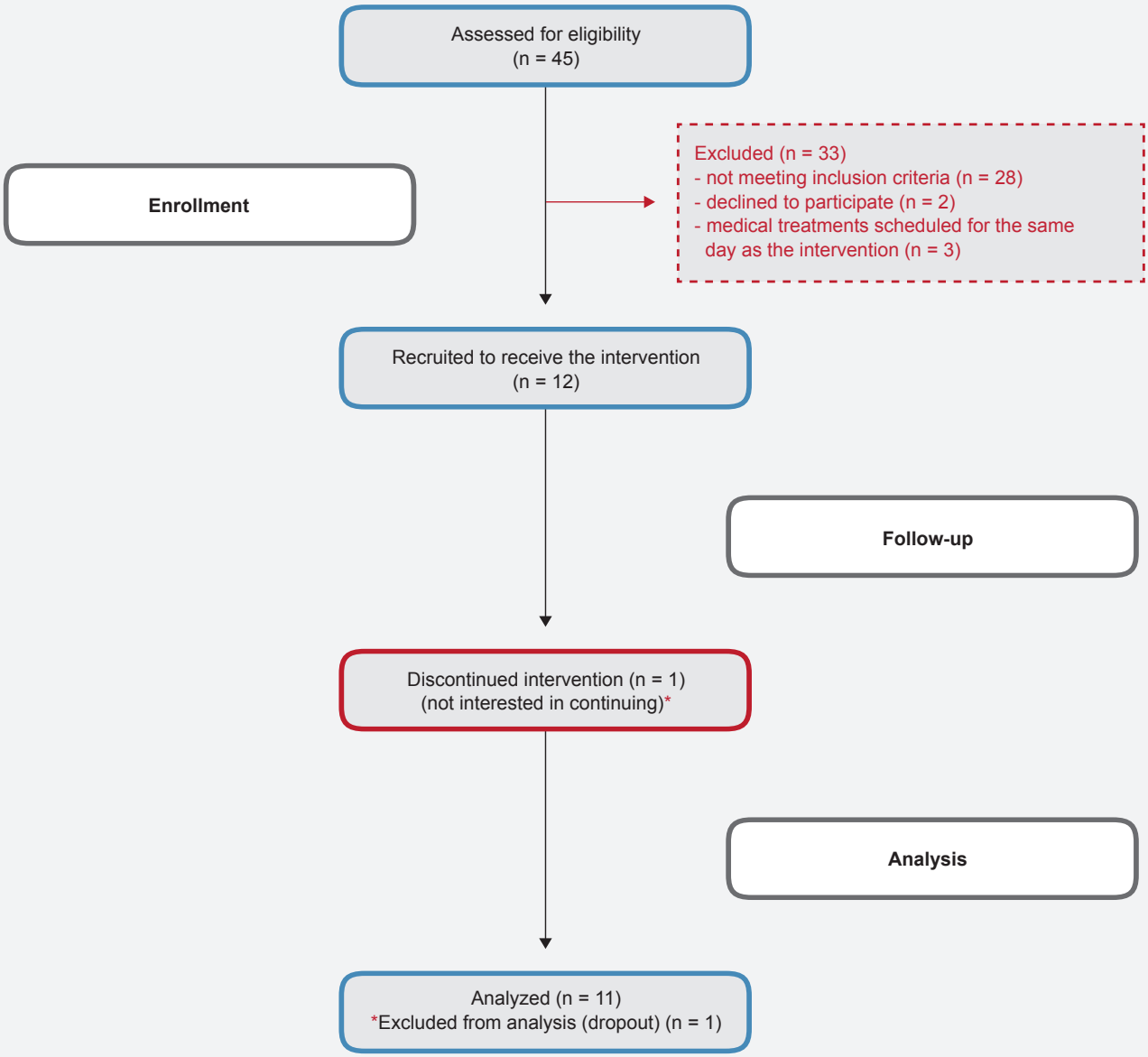

Figure 1 – CONSORT 2010 flow diagram (modified)

Table 1 – Attendance per session

Session	Participants	Therapist	Staff
1	7	1	1
2	10	2	1
3	8	2	1
4	8	1	1
5	8	2	1
6	10	2	1
7	8	2	1
8	9	2	1
9	8	2	1
10	7	2	1

Table 2 – Mean of each VAS item

	Item	Mean	SD	n
BEFORE	**Interest**	27.0	20.3	232
	Response	30.1	21.6	232
	Initiation	23.7	20.6	232
	Involvement	23.8	21.2	232
	Enjoyment	29.7	22.3	232
	Total	13.3	100.0	232
AFTER/DURING	**Interest**	58.3	29.6	225
	Response	63.8	29.7	225
	Initiation	57.1	31.6	225
	Involvement	58.3	31.9	225
	Enjoyment	64.3	31.0	225
	Total	30.5	147.4	225

Table 3 – Inter-rater reliability and test-retest reliability (original data)

	BEFORE		DURING	
	Inter-rater ICC (n = 45) IC	**Test-retest** ICC (n = 94) IC	**Inter-rater** ICC (n = 62) IC	**Test-retest** ICC (n = 119) IC
Interest	0.283* (0.015; 0.519)	0.470** (0.297; 0.612)	0.559** (0.214; 0.751)	0.628** (0.507; 0.725)
Response	0.263* (-0.009; 0.505)	0.464** (0.291; 0.608)	0.477** (0.25; 0.652)	0.593** (0.464; 0.697)
Initiation	0.236 (-0.054; 0.489)	0.361** (0.174; 0.524)	0.694** (0.538; 0.803)	0.742** (0.65; 0.813)
Involvement	0.211* (-0.049; 0.457)	0.592** (0.445; 0.709)	0.528** (0.077; 0.756)	0.700** (0.596; 0.781)
Satisfaction	0.161 (-0.108; 0.417)	0.432** (0.254; 0.582)	0.583** (0.241; 0.767)	0.591** (0.462; 0.696)

ICC case 2 model: two-way random ANOVA, absolute agreement, single measures; CI: 95% confidence interval
*$p < 0.05$; **$p < 0.001$

Table 4 – Factor loadings (principal component analysis) – test 1 (n = 115)

	Interest	Response	Initiation	Involvement	Satisfaction
Interest	1				
Response	0.882**	1			
Initiation	0.785**	0.773**	1		
Involvement	0.848**	0.819**	0.787**	1	
Satisfaction	0.847**	0.860**	0.763**	0.832**	1

Identical results for PCA and PAF
**$p < 0.001$ (two tailed)

Table 5 – Factor loadings (principal component analysis) – test 2 (n = 10)

	Interest	Response	Initiation	Involvement	Satisfaction
Interest	1				
Response	0.984**	1			
Initiation	0.942**	0.923**	1		
Involvement	0.959**	0.961**	0.968**	1	
Satisfaction	0.953**	0.950**	0.917**	0.918**	1

**$p < 0.001$ (two tailed)