Music Therapy Assessment Tool for Awareness in Disorders of Consciousness (MATADOC): Reliability and Validity of a Measure to Assess Awareness in Patients with Disorders of Consciousness

Wendy L. Magee, PhD

Boyer College of Music and Dance, Temple University, Philadelphia, PA, and Royal Hospital for Neuro-Disability, London

Richard J. Siegert, PhD

School of Rehabilitation and Occupation Studies and School of Public Health and Psychosocial Studies, Auckland University of Technology, Auckland, New Zealand

Steve M. Taylor, MSc (hons)

Department of Biostatistics and Epidemiology, Auckland University of Technology, Auckland, New Zealand

Barbara A. Daveson, PhD

King's College London, Cicely Saunders Institute, Department of Palliative Care, Policy, and Rehabilitation

Gemma Lenton-Smith, MA Music Therapy

Ealing Music Therapy, London

Background: Prolonged Disorders of Consciousness (PDOC) describes a population where a consciousness disorder has persisted for at least four weeks post injury but is still under investigation. Complex motor, sensory, communication, and cognitive impairments cause challenges with diagnosis, assessment, and intervention planning. Developing sensitive, reliable, and valid measures is a central concern. The auditory modality is the most sensitive for identifying awareness; however, the current standardized behavioral measures fail to provide adequate screening and measurement of auditory responsiveness. The Music Therapy Assessment Tool

Address correspondence concerning this article to Wendy Magee, PhD, Temple University Presser Hall, 2001 North 13th Street, Philadelphia, PA 19122. Phone: 215-204-1982. E-mail: wmagee@temple.edu

for Awareness in Disorders of Consciousness (MATADOC) is a recently standardized measure for assessment with PDOC; however, psychometric values for two of its subscales require examination.

Objective: *To determine the measurement characteristics and properties of the MATADOC subscales two and three.*

Methods: *In a convenience sample of 21participants with PDOC, a prospective repeated measures study examined inter-rater reliability (IRR) and test-retest reliability (TRR) for both subscales and internal consistency of subscale two.*

Results: *Overall, the items from the MATADOC subscales two and three demonstrated good agreement across and within assessors, with some variability on two identified items.*

Conclusions: *The MATADOC is a standardized measure for assessment of auditory responsiveness in PDOC. Psychometric limitations for the two identified items may have resulted from variations in music therapist clinical experience and training, leading to variations in the administration and interpretation of PDOC patient responses to these two MATADOC assessment items. Although its psychometric properties could be improved, the MATADOC's clinimetric properties make it a valuable assessment to guide clinical work for patients with PDOC.*

Keywords: *music therapy; disorders of consciousness; brain injuries; assessment; vegetative state; minimally conscious state*

———————————————

Introduction

Disorders of Consciousness (DOC) describes a continuum of acquired conditions that stem from acquired profound brain injury. Three primary conditions can be categorized under DOC. Coma represents a state of unarousable unresponsiveness in which there is no evidence of self or environmental awareness (Plum & Posner, 1983). It is usually a temporary phase that progresses to some level of consciousness. Vegetative State (VS) is characterized by spontaneous arousal and sleep-wake cycles but with no evidence of awareness of self or environment, no interaction with others, and no sustained, reproducible, purposeful, or voluntary behavioral responses to sensory stimuli (Jennett & Plum, 1972). Some nerve and spinal reflex responses are preserved, presenting as spontaneous, non-purposeful movements that are not goal directed. Progression from VS leads to Minimally Conscious State

(MCS), where although consciousness remains severely diminished, the person demonstrates minimal but definite behavioral evidence of self or the environment (Giacino et al., 2002). People in MCS are able to demonstrate visual pursuit of moving objects, orient away from noxious stimuli, and may be able to follow simple one-step verbal commands and manipulate objects (Schnakers & Majerus, 2012), although responses typically remain inconsistent.

The nomenclature of "VS" and "MCS" is the most widely accepted at the current time, although there is a call to revise "Vegetative State," given the negative connotations in the English language with the word "vegetable," to the less emotive "Unresponsive Wakefulness Syndrome" (Laureys et al., 2010). Several subcategories of MCS have been proposed to discriminate between people who can demonstrate "higher-level" responses (e.g., command following; communication of "yes" or "no") and those who show less sophisticated or a more limited repertoire of behaviors (e.g., visual tracking and/or orienting away from noxious stimuli only) (Bruno, Vanhaudenhuyse, Thibaut, Moonen, & Laureys, 2011). Descriptors such as MCS+ and MCS− offer sensitive differentiations within MCS and thus make sense to clinicians working with such complex patients who witness subtle variations in patient presentation. However, the revised terminologies are not yet recommended by authoritative sources. The term "Emerging" (i.e., from DOC) describes individuals who show purposeful, reproducible behaviors that are related to the external environment and are more complex in nature. Most recently, DOC has been revised to "Prolonged Disorders of Consciousness" (PDOC) to describe a DOC that has persisted for at least four weeks post injury but is still under investigation (Royal College of Physicians, 2013).

The incidence of PDOC is difficult to determine accurately; however, recent estimates stand at 46 per million in the United States and 14 per million in the UK (Schnakers & Majerus, 2012). The severity of injury and the cost of hospitalization can cause questions about end-of-life (Schnakers & Majerus, 2012). High-profile cases, such as that of Terry Schiavo in 2005, raised awareness that the consequences of a diagnosis with this population can be far reaching: a diagnosis of "VS" as opposed to MCS in that case resulted in the decision to withdraw her feeding tube followed by her death (Schindler v. Schiavo, 2001).

Diagnosis can therefore be seen to influence appropriate care planning and end-of-life decisions alike. Correct diagnosis is

important to enable optimal decision-making for disability management and future care planning, as well as to provide sufficient family support and to make the best decisions regarding the continuation of life-preserving interventions (Magee, 2007). A diagnosis of "MCS" or "Emerging" requires the person with PDOC to demonstrate that he/she has awareness of his/her environment and is able to respond purposefully and differently to contrasting stimuli.

Issues in Diagnosis of PDOC

Diagnosis in PDOC remains a complicated matter (Gill-Thwaites, 2006) despite being a primary concern and aim of clinical intervention, given the unacceptably high rates of misdiagnosis, which are estimated at between 30 and 40% (Hirschberg & Giacino, 2011). As consciousness cannot be directly observed, the assessment of populations with PDOC relies on a number of factors, including rigorous methods of assessment, practitioner experience and skill, and stability of the patient's status (Seel et al., 2010). Behavioral assessment of the PDOC patient relies on careful observations of the person's behaviors at rest (when there is no stimulation), and comparing these with behaviors that occur during the presentation of carefully controlled stimuli. As spontaneous and non-purposeful behaviors are typical of the population, repeated measures are required in order to detect patterns of responsiveness that might indicate awareness.

Functional MRI scanning may provide greater objectivity in measures of awareness (Coleman et al., 2009). However, uncertainty about the accuracy of diagnoses resulting from such tests due to technical difficulties confirms behavioral assessments as the preferred method (Laureys & Schiff, 2012). Thus, much of the research with PDOC populations focuses on developing standardized measures for diagnostic assessment, outcome prediction, interdisciplinary treatment planning and projection of future needs, and monitoring treatment effectiveness (Giacino, Kalmar, & Whyte, 2004).

Developing reliable and valid behavioral measures for assessment is challenging due to the severity of disability typical of the population: a combination of motor, sensory, and cognitive impairments can mask residual functioning, thus risking misdiagnosis. The auditory modality has been the focus of greater empirical inquiry due to evidence that it is the more sensitive modality for

identifying awareness (Gill-Thwaites & Munday, 1999; Owen et al., 2005), provided by clinical reports of persons with DOC who have demonstrated cognitive capacity through the auditory modality even in the absence of movement and language (Giacino et al., 2009; Owen et al., 2006). Despite the increased evidence for using auditory stimuli to assess awareness, the existing standardized measures for DOC populations fail to address auditory responsiveness adequately (Lichtensztejn, Macchi, & Lischinsky, 2014; Magee, Siegert, Daveson, Lenton-Smith, & Taylor, 2014). The recommended evidence-based measures reduce auditory stimuli to verbal material or stimuli that cause a startle response (e.g., blowing a whistle, clapping hands, banging wooden blocks) or involve calling the patient's name (Seel et al., 2010).

Music Used for Diagnosis in PDOC

Music is receiving increasing attention as a stimulus for assessment of awareness given its properties of being a non-language-based stimulus within the auditory modality that has emotional saliency (Boyle & Greer, 1983; Formisano et al., 2001; Jones et al., 2000; Lichtensztejn et al., 2014; Magee, 2005, 2007; Magee et al., 2014; O'Kelly et al., 2013; Okumura et al., 2014; Perrin et al., 2006; Puggina, da Silva, & Santos, 2011; Verger et al., 2014; Verville et al., 2012; Wilson, Cranny, & Andrews, 1992). Neurophysiological studies have revealed that music can promote behavioral responses indicative of arousal and attention in PDOC patients, and most importantly, that significant cortical and autonomic nervous system activity suggest discriminative ability rather than reflex responses (O'Kelly et al., 2013). Furthermore, music may not just be a useful tool for evaluation of consciousness, but may stimulate brain activation that can predict patient improvement (Okumura et al., 2014).

Despite its promise as a diagnostic and prognostic tool, robust measures and sensitive protocols are needed for using music-based assessments and interventions in PDOC. Some music therapists working with PDOC populations currently use standardized neuropsychological measures to assess patient improvements in response to music therapy (Lichtensztejn et al., 2014). However, such measures do not provide adequate protocols for testing auditory responsiveness and are not sufficiently specific about responses to auditory stimuli (Lichtensztejn et al., 2014). This

misses an opportunity to formulate patient-centered recommendations for individuals' rehabilitation and care. Drawing on the evidence underpinning music as a prognostic and diagnostic medium in PDOC, a measure for assessing awareness in PDOC that involves the manipulation of music-based stimuli in the auditory modality was developed to provide a more sensitive assessment of auditory responsiveness, the Music Therapy Assessment Tool for Awareness in Disorders of Consciousness (MATADOC).

Addressing Evidence-Based Requirements for Measures for DOC

A number of recommendations for evidence-based behavioral measures utilized with DOC populations have been specified (Seel et al., 2010) and were addressed within the current study. First, we provided specialist training, which included a comprehensive manual and clearly defined protocol (Magee, Lenton-Smith, & Daveson, 2012) to enable consistent administration and scoring of items, and to provide interpretive guidelines that fit the Aspen Workgroup consensus-based diagnostic classifications (Giacino et al., 2002). Second, to address content validity, behaviors outlined in the MATADOC are representative of VS, MCS, and emerging behaviors as outlined in the Aspen Workgroup criteria (Giacino et al., 2002). Third, we addressed factors for establishing reliability by ensuring that assessors had sufficient clinical experience (0.25–19 years, mean = 3.5 years) with PDOC populations, and all assessors were trained to a specified level of competency in using the MATADOC. In addition, when examining test-retest reliability, change in patient status between scale administrations was minimized through the study's design by using video records. The use of video records allowed us to address the characteristic progression in state of consciousness for PDOC populations. Fourth, criterion validity was examined through a comparison of the MATADOC to a standardized reference standard, the SMART (Gill-Thwaites & Munday, 1999), which has been evaluated as having good content validity and may be used to assess DOC with moderate reservations (Seel et al., 2010). Fifth, construct validity based on the Rasch model has been examined for the MATADOC Principal Subscale, which is the only subscale that carries diagnostic utility (Magee et al., 2014). MATADOC subscales two and three have utility primarily for goal setting and intervention planning.

The Rasch analysis indicated that all five items of the Principal Subscale demonstrated good fit to the Rasch model for the individual items and the overall model fit. Both principal component and Rasch analyses of the Principal Subscale, therefore, demonstrate a robust unidimensional and homogeneous subscale for assessing awareness in patients with PDOC. The final criterion, that a scale should predict outcome, remains unknown at the current time.

Purpose of the Current Study

Given a previous study to explore the reliability and dimensionality of the MATADOC Principal Subscale (Magee et al., 2014), the purpose of this study was

   i) to examine the reliability (interrater and test-retest) of subscales two and three of the MATADOC, including those items that previously had not been examined; and

  ii) to explore the internal consistency of the two items in the second subscale that use binary rating for a group of behaviors given their clinical rather than diagnostic utility.

Method

Design

We employed a prospective study using repeated measures to test the reliability of subscales two and three and the internal consistency of the second subscale.

Description and Purpose of the MATADOC

The MATADOC is a 14-item measure that was developed and refined over a 17-year period for assessment of responsiveness in adults with PDOC. It is therefore clinically derived with an emphasis on its clinical utility within interdisciplinary assessment and treatment of adults with PDOC following acquired profound brain injury. Its purpose is to provide a rigorous and detailed assessment of auditory responsiveness in PDOC patients, given the evidence that the auditory modality is more sensitive to detecting awareness in this population (Gill-Thwaites & Munday, 1999; Owen et al., 2005). In addition to the auditory modality, it also examines other behavioral domains that are tested as standard using a protocolized set of procedures (Magee et al., 2012). The MATADOC's overall

purpose is to demonstrate whether the person with PDOC is aware by measuring whether he or she responds differently and purposefully to contrasting stimuli and to guide intervention planning accordingly.

The MATADOC has three subscales that each serve different purposes and are described here briefly, followed by a more detailed description of each subscale (see Table 3). The "Principal Subscale" tests and reports on responsiveness across several behavioral domains that are included in all DOC measures and contribute to a diagnosis. The second subscale, "Musical Parameter and Behavioral Response Type," reports on musical behaviors to inform music therapy intervention planning. The third subscale, "Clinical Information to Inform Goal Setting and Clinical Care," assesses responses that can be considered indicative of cognitively mediated responses.

Principal Subscale: Essential Categories.

The Principal Subscale has a total of five items, and assesses the following five areas: responses to visual stimuli, responses to auditory stimuli, awareness of musical stimuli, responses to verbal commands, and arousal. Two items within this subscale assess auditory responsiveness: one rates responses to auditory stimuli more generally, for example to voices, to single musical sounds, as well as to more complex musical stimuli. The second auditory item rates responses to musical stimuli more specifically, such as songs sung using the patient's name and familiar music that is known to be personally meaningful to the patient. The Principal Subscale is most closely aligned to other validated assessment measures for PDOC, given that it tests the patient across the auditory, visual, and communication domains as well as wakefulness (Seel et al., 2010). These are considered "essential" domains to assess in DOC measures. The psychometric properties and reliability of this subscale have been reported previously as having good interrater reliability (mean = 0.83, sd = 0.11) and good test-retest reliability (mean = 0.82, sd = 0.05) (Magee et al., 2014). Internal consistency for the five-item Principal Subscale was rated as good (α = 0.76) for a four–six-item scale for DOC populations (Seel et al., 2010) with a strong first principal component. Rasch analysis confirmed these impressions of a reliable, unidimensional, and homogeneous scale.

The MATADOC Principal Subscale has been compared with an external reference standard to test its concurrent validity. The Sensory Modality Assessment and Rehabilitation Technique, or SMART (Gill-Thwaites, 1997), is an established and standardized global measure for PDOC that assesses responsiveness across sensory modalities (Gill-Thwaites & Munday, 1999). Excellent agreement (100%) was observed between MATADOC and SMART outcomes for diagnosis of awareness states, with Cohen's $\kappa = 1$ providing strong evidence of validity and Fisher's exact test demonstrating a strong association ($p < 0.001$) (Magee et al., 2014). However, contrasting sensitivities between the measures have been highlighted, with MATADOC having a higher sensitivity within the auditory and visual domains relative to SMART, and SMART having a higher sensitivity in the motor domain (O'Kelly & Magee, 2013). These findings indicated that both measures have contrasting sensitivities, providing data predictive of awareness from mutually exclusive behavioral domains: SMART and MATADOC may elicit different levels of response in comparable domains. The musical response items in MATADOC provide distinctive behavioral data that may contribute to prediction of awareness. These findings highlight the importance of concurrent use of varied measures to provide a comprehensive multisensory assessment of PDOC patients, supporting the recommendations of Giacino et al. (2002). Although the MATADOC has demonstrated excellent concurrent validity with the SMART, its predictive validity remains untested.

Second Subscale: Musical Parameter and Behavioral Response Type.

The second subscale has a total of two items, and assesses two areas (Behavioral Response to Music; Musical Response) that inform intervention planning. Item 6: Behavioral Response to Music rates behaviors contingent to musical stimuli across six sub-items, such as wakefulness, physical movement, facial gesture, respiration rate and eye direction, and vocalization. Item 7: Musical Response has seven sub-items that rate patient responsiveness to specific musical parameters (pulse/rhythm, melody/pitch, timbre, dynamics/intensity, form, tempo, and musical mood). The purpose of this subscale is to provide detailed information about responsiveness to specific components of the auditory environment

and any behaviors that are contingent to stimuli. For example, responses to manipulation of volume can provide important information for managing the patient's auditory environment; contingent responses to manipulation of timbre (e.g., change of instrument) or musical form (e.g., pauses in the music) can provide evidence of purposeful response that can contribute to a diagnosis of awareness. Detailed information of this type helps both tailor individualized music therapy intervention optimally suited to an individual patient's responsiveness as well as provide information to assist carers in structuring the patient's auditory environment. This is important for people who are fully dependent on others for all aspects of care.

Third Subscale: Clinical Information to Inform Goal Setting and Clinical Care.

The third subscale has a total of seven items, and assesses motor, communication, and emotional behaviors. This subscale serves to inform goal setting and clinical care. Two items examine communicative behaviors drawing on musical behaviors: nonverbal communication and vocalization. Thus, six items of the entire MATADOC measure focus on auditory responsiveness. The seven items of the third subscale extend beyond the essential behaviors that are assessed in PDOC (e.g., visual responsiveness, wakefulness) and that are covered in the Principal Subscale. All of the items in the third subscale rate behaviors that are associated with cognitively mediated behaviors, for example Item 10: Choice-Making or Item 13: Intentional Behaviour. Thus, the items are well suited for procedures that develop specific patient strengths as part of interdisciplinary goal-oriented rehabilitation.

The MATADOC Protocol

The MATADOC protocol uses a wide range of musical stimuli, including isolated single auditory stimuli (i.e., a single pitch on a single instrumental timbre), complex musical sounds (e.g., the strum of a guitar involving multiple pitches on a single instrumental timbre; singing a familiar song with keyboard accompaniment enabling the manipulation of instrumental timbre, pitch, harmony, and volume), and musical activities to measure a number of functional behaviors across the motor, communication, visual, and

auditory domains (e.g., goal-directed movement to make a sound on an instrument; use of vocalization within a song) (Magee et al., 2012).

At its simplest, the MATADOC protocol involves a minimum of five procedures that aim to elicit patient responses. Depending on responsiveness, the five procedures can be extended to assess and develop, where possible, more complex behaviors such as pragmatic and functional communication, motor responses, or cognitively mediated functions such as choice-making and attention to task. All procedures use music or music-related stimuli. The music used is predominantly presented live, although recorded music can be used within one procedure for genres that are known to be personally meaningful to the patient but are difficult to recreate convincingly by an assessor (e.g., opera, hip hop, world). The music used varies between novel unfamiliar music and salient music known to have personal meaning to the patient. Personally meaningful music is determined through a prior interview with family. The assessment protocol is followed and administered over the course of four individual assessment sessions, lasting between 15 and 30 minutes each depending on patient tolerance, within a ten-day period. A short behavioral observation period before and after each assessment session allows for data collection of "at-rest" behaviors.

Scoring the MATADOC

The MATADOC provides three scores that contribute to a diagnosis (see Table 3). After rating the observed behaviors across 14 items, the assessor converts the ratings into a numerical score for the Principal and second subscales. The third subscale converts ratings of each of its items to an associated behavioral category of "VS, MCS, or Emerging.".

The score produced by the Principal Subscale has diagnostic utility (Magee et al., 2014). The score provided by subscale two primarily has clinical utility. It uses binary rating (yes/no) to identify which musical components elicit observed behavioral responses, thereby assisting the music therapist with planning intervention. The binary scoring used in subscale two differentiates it from the other subscales. The scored outcomes in subscale three (VS, MCS, or Emerging for each item) support the diagnostic score provided by the Principal Subscale.

Participants

A convenience sample was recruited from a specialist unit for adults with PDOC who were admitted to the facility for the purposes of gaining an accurate diagnosis through interdisciplinary assessment. Participants were required to have no confirmed diagnosis of VS, MCS, or Emerging at the time of recruitment. Participants were required to be medically stable and be aged between 16 and 70 years of age. Participants with known pre-morbid hearing impairments, a previous diagnosis of musicogenic epilepsy, or who emerged from DOC during recruitment or assessment were excluded. Mental capacity assessments were completed before recruitment, and standard procedures regarding the recruitment of patients lacking capacity were followed. Ethics was gained from the UK National Research Ethics Service (05/Q0406/185). All participants received standard interdisciplinary rehabilitation assessment and intervention during the study.

Procedures for Data Collection

Data were collected for each participant during four assessment sessions within a 10-day period. The MATADOC protocol that has previously been described (Magee et al., 2012) was followed in all assessment sessions. Two independent assessors were involved in every assessment session (Assessors A and B). Assessor A delivered the protocol and rated behaviors, and Assessor B observed and rated behaviors, thus the same live assessment session was rated simultaneously by two independent assessors. A paper form of the MATADOC assessment was completed by each assessor at the end of each assessment session. The four live assessment sessions were video recorded, and the video records were rated again by the same two assessors using paper forms of the MATADOC at a later date for test-retest analysis.

Three-minute observational periods were conducted at rest immediately before and after each MATADOC session in line with best evidence-based practice recommendations for DOC (Seel et al., 2010). Behavioral data collected during these observations did not contribute to scoring. Behavioral observations with no stimulation are essential to determine whether the behaviors observed during musical stimuli are purposeful, as the purpose of DOC measures is to determine whether the person with PDOC responds differently and purposefully to contrasting stimuli.

Procedures to Minimize Bias

All the assessors in this study were credentialed music therapists who were trained in using the MATADOC. Seven assessors in total were involved in data collection, whose experience with PDOC populations ranged from 0.25 to 19 years (mean = 3.5 years). The two assessors involved with each individual participant remained consistent across that individual's entire MATADOC assessment (four live clinical contacts and four video ratings). Assessors were blinded to each other's ratings for interrater analysis. An adequate period of time (between four and 24 weeks) elapsed between the live session and the video observation to minimize the assessor's memory of the previous session. Furthermore, viewing order of the clinical contacts was randomized, providing further interference with any memory of the live MATADOC ratings. All data were anonymized and entered into an EXCEL spreadsheet by a data manager who was independent of this project. Later, data were imported into SPSS for statistical analysis.

Analysis

Interrater and test-retest reliability were assessed using the intraclass correlation coefficient with random effects, commonly referred to as ICC(2), to account for there being multiple raters ($n = 7$). Interrater reliability analysis used only live assessment sessions but included all repeated contacts ($n = 4$ each), which were assumed to be independent. This resulted in a sample size of 168 (21 x 4 x 2) observations for interrater analysis. A similar type of assumption was made for test-retest. Test-retest reliability analysis compared ratings of live and video observations of the same contact undertaken by the same assessor (168 ratings: 21 x 4 x 2). The intra-patient variation across multiple contacts in a PDOC population is sufficient to support the use of this technique (see Figure 1).

The two items contained within the second subscale (MATADOC items 6 and 7) were examined independently for internal consistency. This was necessary, given that these items were structured and scored differently to the items in the Principal and third subscales using binary rating. This subscale warranted special examination, given its primary utility to inform music-specific intervention planning and the arbitrary nature of its numerical scoring: "no" was scored 1, and "yes" was scored 2. Internal consistency of scores

	Rater A: Live rating.	Rater B: Live rating.	Rater A: Video rating.	Rater B: Video rating.
Rater A: Live rating.		Inter-rater reliability. $n =$ 84	Test-retest. $n =$ 84	
Rater B: Live rating.	Inter-rater reliability. $n =$ 84			Test-retest. $n =$ 84

Figure 1.
Matrix of live and video ratings to achieve inter-rater and test-retest ratings.

derived from multiple responses was assessed for items 6 and 7 using Cronbach's alpha, using live observations only (excluding video observations) and assuming the correlations of items do not change over repeated assessments of the same patients. All analyses were conducted using R version 3.1.2 (R Core Team, 2014) and the "psych" package (Revelle, 2014).

Criteria for assessing measures for rehabilitation vary significantly in the literature. Although recommendations propose a rating scheme for interrater and test-retest reliability in DOC scales where coefficients of < 0.70 are unacceptable (Seel et al., 2010), others have stressed that measures for disability should balance traditional psychometric and clinimetric considerations of such measures: criteria have been proposed where coefficients of > 0.40 to < 0.75 are rated as adequately covered but with weakness in some domains or lacking in a noncentral domain (Andresen, 2000). Given the primary purposes of the two subscales under examination being for clinical rather than diagnostic utility, we have adopted the recommendations laid forth by Andresen (2000).

Results

Sample Characteristics

Over a 36-month period, twenty-one participants (11 male; 10 female) were recruited with a mean age of 40.3 years (range = 19–67; SD = 15.65). All participants had profound brain

damage and were completely dependent for all activities of daily living, with profound communication impairments and unknown sensory responsiveness. Etiology of brain injury covered trauma (n = 9), hypoxic/ischaemic (n = 7), haemorrhagic (n = 3), and viral infection (n = 2). The average time since onset was 7.9 months (see Table 1 for total characteristics). MATADOC outcomes for the sample ranged across VS (n = 13) and MCS (n = 8). There were no cases who were diagnosed as Emergent. These diagnoses were in 100% agreement with diagnoses provided by an external reference standard (see Magee et al., 2014).

Reliability

We report in detail here on subscales two and three (items 6–14) of the MATADOC, as IRR and TRT for the Principal Subscale have already been reported (Magee et al., 2014). Calculated intra-class correlation coefficients (ICCs) for items from subscales two and three of the MATADOC produced mixed results. ICCs for

TABLE 1

Patient Cohort Recruited to the Study

Patient number	Gender	Age	Etiology of brain damage	Time since onset (months)
1	M	21	Traumatic	7
2	M	25	Traumatic	8
3	M	60	Hypoxic/ischaemic	5
4	M	23	Traumatic	10
5	M	19	Hypoxic/ischaemic	9
6	F	42	Hypoxic/ischaemic	8
7	F	19	Traumatic	13
8	M	48	Haemorrhagic	5
9	F	27	Viral	15
10	F	37	Traumatic	6
11	F	35	Hypoxic/ischaemic	6
12	F	47	Haemorrhagic	8
13	F	59	Traumatic	5
14	M	36	Hypoxic/ischaemic	6
15	M	65	Traumatic	16
16	F	58	Viral	10
17	M	44	Traumatic	7
18	M	45	Hypoxic/ischaemic	5
19	M	23	Traumatic	6
20	F	67	Hypoxic/ischaemic	5
21	F	46	Haemorrhagic	5

IRR ranged from 0.32 to 0.74 (mean = 0.48) for seven of the nine items, with a mean ICC of 0.41 for the two subscales overall. Item 7: Musical Response and Item 10: Choice-Making had much lower ICCs, 0.14 and 0.17, respectively. Using the criteria proposed by Andresen (2000), Items 8–9 and 11–12 had adequate IRR but Items 7, 10, and 14 had poor IRR, with items 6 and 13 also falling just inside the threshold for poor IRR (0.39).

Test-retest reliability for most of the items fell within the acceptable range (range = 0.55–0.69, mean = 0.61) with exception of Item 7: Musical Response (0.26) (see Table 2). The results indicate acceptable TRR for Items 6 and 8–14. However, Item 7: Musical Response has poor TRR.

Internal Consistency

Items 6 and 7 were examined independently for internal consistency, given that they each have nonhierarchical sub-items with binary rating. They were examined as separate rather than combined items, as they do not examine the same construct. Cronbach's alpha for Item 6: Behavioural Response (with six sub-items) was 0.51; for Item 7: Music Response (seven sub-item), it was 0.28. These results indicate poor internal consistency for Items 6 and 7 (See Table 2).

Discussion

This study aimed to examine the reliability of the two subscales of the MATADOC, "Subscale Two: Musical Parameter and Behavioural Response Type" and "Subscale Three: Clinical Information to Inform Goal Setting and Clinical Care." These subscales were examined separately, as they were believed to hold greater clinical utility rather than psychometric strength. This contrasts with the MATADOC Principal Subscale, which holds diagnostic utility and so was tested for its psychometric properties separately. The three subscales have combined strength as a comprehensive measure that can rate responsiveness to music-based auditory stimuli in PDOC populations. Thus, this study aimed to provide a clearer picture of the MATADOC's overall psychometrics.

Testing for IRR and TRT for subscales two and three of the MATADOC entire measure revealed mixed results. Four of the items had adequate reliability, with a further two reaching just below the

TABLE 2

Descriptive Statistics (N = number of pairs of observations, M = mean score, SD = standard deviation), Internal Consistency (Cronbach's α) of Derived Scores, and Interrater and Test-Retest Reliability (estimated using intraclass correlation coefficients, ICC) for MATADOC Subscales 2 and 3

Variable	N	M	SD	Internal consistency		Interrater reliability		Test-retest reliability	
				α	95% CI	ICC	95% CI	ICC	95% CI
Score Q6, items summed	82	1.40	1.33	0.51	(0.37, 0.65)	0.39	(0.17, 0.54)	0.62	(0.52, 0.70)
Score Q7, items summed	84	0.36	0.67	0.28	(0.10, 0.46)	0.14	(−0.08, 0.34)	0.26	(0.11, 0.39)
Vocalization	84	−0.01	0.83			0.74	(0.63, 0.82)	0.60	(0.49, 0.69)
Nonverbal communication	84	0.37	0.75			0.63	(0.48, 0.74)	0.63	(0.53, 0.72)
Choice-making	84	0.08	0.32			0.17	(−0.04, 0.37)	0.58	(0.47, 0.67)
Motor skills	84	0.58	0.86			0.47	(0.25, 0.59)	0.61	(0.50, 0.69)
Attention to task	84	0.45	0.57			0.43	(0.21, 0.57)	0.61	(0.50, 0.69)
Intentional behaviour	84	0.52	0.83			0.39	(0.18, 0.55)	0.69	(0.60, 0.76)
Emotional response	84	0.57	1.02			0.32	(0.12, 0.50)	0.55	(0.43, 0.64)

TABLE 3

MATADOC Subscales, Items and Their Purpose, and Score Derivations

Subscale title/Items	Purpose	Method of score derivation
	Principal Subscale (Essential categories)	
1: Responses to visual stimuli	Rates responses to visual stimuli presented without an accompanying auditory component.	Ratings are converted to a score of 0–2, where 0 is indicative of a "VS" level response; 1 indicative of an "MCS" level response; and, 2 indicative of an "Emerging" level response.
2: Responses to auditory stimuli	Rates responses to auditory stimuli presented without visual stimuli for assessing all evidence of residual hearing including startle responses.	
3: Awareness of musical stimuli	Rates behavioral and musical responses that provide evidence of the patient's awareness of the music (live or recorded).	The summed total of the five items' scores are linked to a diagnostic category:
4: Response to verbal commands	Rates responses to one-step verbal commands.	a) 0–3 = VS b) 4–7 = MCS
5: Arousal	Rates arousal across the session demonstrate by eyes open and being alert.	c) ≥ 8 = Emerging
	Second subscale: Musical parameter and behavioraul response type	
6: Behavioural response to music	Rates all observable cognitively mediated behaviors at the initial presentation of musical stimulation or throughout other periods within the assessment session. Records differential responses, i.e., behaviors which occur contingent to musical stimuli in 6 sub-items, each of which is rated independently.	Each sub-item has a binary rating: "No" score = 1; "Yes" score = 2. Total score range for subscale: 13–26.
7: Musical response	Rates responses to musical parameters in 7 sub-items, each of which is rated independently. Useful for collecting information to inform subsequent music therapy interventions specifically tailored to meet the patient's responses.	Sum of scores tracks developments or losses in overall responsiveness. Holds no diagnostic value but assists with intervention planning for use of musical stimuli.

TABLE 3

Continued

Subscale title/Items	Purpose	Method of score derivation
	Third subscale: Clinical information to inform goal setting and clinical care	
8: Vocalization	Documents the patient's vocal activity (including reflexive behaviors and cognitively mediated responses).	Item ratings are converted to scores that are categorized as Absent/VS, Developing/MCS, or Strength/Emerging responses.
9: Nonverbal communication	Rates observed responses pertaining to nonverbal social communication responses.	Holds no diagnostic value but
10: Choice-making	Rates the ability to discriminate between a forced choice of two alternatives presented using the communication strategy used by the treatment team.	a) supports the diagnosis formed from the Principal Subscale;
11: Motor skills	Rates observed physical responses to musical objects, sounds, and music presented.	b) assists with uni/interdisciplinary goal setting; and
12: Attention to task	Rates volitional behaviors denoting attention (focused or sustained) during procedures involving musical stimuli, be they auditory (e.g., musical sounds) or combined auditory/visual.	c) assists with monitoring patient responsiveness against the goals that have been set.
13: Intentional behaviour	Rates observed goal-oriented behaviors.	
14: Emotional response	Rates emotionally expressive behaviors that are contingent to musical stimuli including facial gestures, tears, and vocalizations.	

threshold of "adequate" for IRR. TRT reliability was stronger overall for nearly all items. Item 10: Choice-Making had poor IRR, and Item 7: Musical Response had poor IRR and TRT. Furthermore, Item 7 and Item 6: Behavioural Response have poor internal consistency. This contrasts with the MATADOC Principal Subscale that as a five-item scale has good IRR and TRT and good internal consistency (Magee et al., 2014).

Given the poor results for "Item 7: Musical Response," one option is to remove this item from the MATADOC. However, as this is the primary item for rating music-specific parameters that have elicited responsiveness in this complex population, its removal would be shortsighted. Therefore, examining underlying reasons for its poor IRR/TRT warrants further explanation. Several factors that may have contributed to poor IRR/TRT for Item 7 include limited sensitivity to change due to the scoring system, assessor experience, and MATADOC training.

First, the sub-items are added into an arbitrary score that has no real value: the "score" offers a numerical figure based on the clinician's simple behavioral ratings of "present" or "absent" behaviors. These ratings can be elaborated in a useful manner to guide decisions in tailoring the intervention for the next assessment session. So, it is stressed that the scores derived from Item 7 do not contribute to diagnosis, but remain relevant for planning intervention.

The purpose of Item 7: Musical Response is to guide clinical decision-making for tailoring intervention to an individual's specific responses to musical stimuli. Breaking down auditory responsiveness in this way is important to enhance the sensitivity of the MATADOC assessment; not only does it assist the music therapist to tailor intervention, but it also guides recommendations for managing the auditory environment (e.g., in ward environments) and can contribute to global communication strategies. These contributions are important for the heavily dependent patient with profound communication impairment who is reliant on others to manage her/his auditory environment (e.g., whether or not to play music, television, etc., or have a period of no stimulation or silence). However, reducing musical experiences to musical parameters (e.g. "tempo"; "timbre") for rating in this way may be an underlying problem for the psychometric properties of this item. That is, the variance between the sub-items may just be too small. Thus, Item 7 is weak psychometrically for reasons of scoring

but holds clinical utility, as it focuses the therapist's attention to the musical parameters relating to auditory responsiveness in the PDOC patient.

Second, it could be argued that this item may reveal variance in the assessors' skills. Rating responses on this item requires the therapist/assessor to make decisions as to whether the observed behaviors are related specifically to the music being played in the session by the therapist/assessor. For example, the therapist assesses whether motor movements observed are at a similar tempo to music being played or reflect the pulse of the music; or whether client-initiated vocal sounds are at the pitch of the music being played. Despite training to a recognized level of competency, the skills required to assess patient responses in this complex population are highly variable between assessors. Within the music therapy profession, there is considerable variation as to music-centered practices, influenced by factors such as the model in which the therapist trained, the clinical population worked with, and the care setting worked in. Similarly, there was little consistency within the research team in this study that comprised seven therapists from six different training backgrounds. In reality, some therapists will make more use of Item 7 than others, and use it differently from others, for example the degree to which a therapist makes links between (musical) cause and (behavioral outcome) effect. However, removing the item altogether prevents a music therapist from making a finely tuned assessment of the patient's auditory responsiveness to musical parameters. In essence, it is Item 7: Musical Response that makes the MATADOC a music-based assessment.

Third, poor IRR and TRT suggest that the training for therapists to deliver and rate this item is currently inadequate and needs further attention. Feedback from follow-up MATADOC training sessions has revealed that trainees are not consistent in their understanding of how to deliver protocol procedures relating to this item, nor of how to assess responses. This is now being addressed in MATADOC training, using illustrative online videos and greater explanation of how to assess and rate Item 7 and its relevance for patient care. Thus, despite the item's poor psychometric properties, it has clinimetric value.

Finally, this study involved a relatively small sample of patients who do not demonstrate much change over time. Consequently, this sample may not have provided enough variance to properly

evaluate the psychometric quality of this item. A sample that included patients with a greater variance of consciousness (i.e., including people diagnosed as "Emerging") may show Item 7 in a better light or at least enable improved evaluation.

The internal consistency of both Items 6 and 7 suggests that the within-item constructs are poorly related. Given the range of constructs included in each item, this seems accurate. This is not to suggest that the items are not useful or unimportant, but simply that no importance should be given to the arbitrary single total scores they produce, as this could be analogous to adding apples and oranges and pears.

Similarly, Item 10: Choice-Making was revealed to be a weaker item in terms of interrater reliability. Item 10 rates abilities that concern developing communication skills and using these to communicate preference (e.g., for an instrument; for a song; for a favorite music performer). Thus, it is clear that this item is appropriate only for MCS or Emergent patients. The variability we found can be explained by the difficulty in using a single systematic procedure to assess for this behavioral domain. "Choice-Making" in this patient group relies on the patient's developing communication skills. Typically with this patient group, a variety of behaviors emerge to indicate "communication" in the period of the patient's recovery when a communication strategy is developing. The complexity of patient response is highly variable (e.g., eye gaze versus gesture for "yes") and often inconsistent. In addition, variability in assessors' ability and confidence to assess this skill may have contributed to this item being weaker in terms of IRR. As with Item 7, this has now been addressed in training and is given greater explication. In reality, this is an advanced clinical practice skill for professionals working in PDOC and typically relies on learning from interdisciplinary colleagues with specialist skills.

Conclusion

Overall, the results need to be considered in light of this measure being a work in progress. The entire MATADOC offers clinicians a useful measure for documenting behaviors in a consistent manner. It demonstrates sensitivity through its capacity to rate behaviors in small enough increments to reflect change, a necessary function for evaluating responsiveness in minimally responsive populations.

It provides a tool for assessment and evaluation of intervention, and thus has potential for demonstrating patient change over time. It has clinical utility due to its ease of use once trained, its adaptability for a range of music therapy practices, and its ability to provide information that is essential to diagnosis of the PDOC patient.

In a previous study, we established that the Preliminary Subscale has diagnostic utility, including good IRR, TRT, and internal consistency (Magee et al., 2014). In this study, we established that subscales two and three have IRR and TRT to a degree that suggests the MATADOC's utility for guiding clinical intervention. To improve MATADOC implementation, training now gives greater priority to three areas: 1) presenting procedures for musical responsiveness assessment, 2) how to assess behaviors that can be deemed as musical responsiveness, and 3) training assessors in advanced skills to assess the patient's communication strategies. Future research should include a larger and more varied sample that includes VS, MCS, and Emerging patients to provide improved examination of the second subscale.

References

Andresen, E. M. (2000). Criteria for assessing the tools of disability outcomes research. *Archives of Physical Medicine and Rehabilitation*, *81*, S15–S20.

Boyle, M. E., & Greer, R. D. (1983). Operant procedures and the comatose patient. *Journal of Applied Behavior Analysis*, *16*(1), 3–12.

Bruno, M., Vanhaudenhuyse, A., Thibaut, A., Moonen, G., & Laureys, S. (2011). From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: Recent advances in our understanding of disorders of consciousness. Journal of Neurology, *258*(7), 1373–1384. doi:10.1007/s00415-011-6114-x

Coleman, M. R., Davis, M. H., Rodd, J. M., Robson, T., Ali, A., Owen, A. M., & Pickard, J. D. (2009). Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. *Brain: A Journal of Neurology, 132*, 2541–2552. doi:10.1093/brain/awp183

Formisano, R., Vinicola, V., Penta, F., Matteis, M., Brunelli, S., & Weckel, J. W. (2001). Active music therapy in the rehabilitation of severe brain injured patients during coma recovery. *Annali Dell'Istituto Superiore Di Sanità, 37*(4), 627–630.

Giacino, J. T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D. I., … Zasler, N. D. (2002). The minimally conscious state: Definition and diagnostic criteria. *Neurology, 58*(3), 349–353.

Giacino, J. T., Kalmar, K., & Whyte, J. (2004). The JFK coma recovery scale-revised: Measurement characteristics and diagnostic utility. *Archives of Physical Medicine and Rehabilitation, 85*(12), 2020–2029.

Giacino, J. T., Schnakers, C., Rodriguez-Moreno, D., Kalmar, K., Schiff, N., & Hirsch, J. (2009). Behavioral assessment in patients with disorders of consciousness: Gold standard or fool's gold? *Progress in Brain Research, 117*, 33–47.

Gill-Thwaites, H. (1997). The sensory modality assessment rehabilitation technique—A tool for assessment and treatment of patients with severe brain injury in a vegetative state. *Brain Injury, 11*(10), 723–734. doi:10.1080/026990597123098

Gill-Thwaites, H. (2006). Lotteries, loopholes, and luck: Misdiagnosis in the vegetative state patient. *Brain Injury, 20*(13–14), 1321–1328.

Gill-Thwaites, H., & Munday, R. (1999). The sensory modality assessment and rehabilitation technique (SMART): A comprehensive integrated assessment and treatment protocol for the vegetative state and minimally responsive patient. *Neuropsychological Rehabilitation, 9*(3–4), 305–320. doi:10.1080/096020199389392

Hirschberg, R., & Giacino, J. T. (2011). The vegetative and minimally conscious states: Diagnosis, prognosis, and treatment. *Neurologic Clinics, 29*(4), 773–786. doi:10.1016/j.ncl.2011.07.009

Jennett, B., & Plum, F. (1972). Persistent vegetative state after brain damage: A syndrome in search of a name. *Lancet, 1*(7753), 734–737.

Jones, S. J., Pato, M. V., Sprague, L., Stokes, M., Munday, R., & Haque, N. (2000). Auditory evoked potentials to spectro-temporal modulation of complex tones in normal subjects and patients with severe brain injury. *Brain, 123*(5), 1007–1016.

Laureys, S., Celesia, G. G., Cohadon, F., Lavrijsen, J., León-Carrión, J., Sannita, W. G., … the European Task Force on Disorders of Consciousness. (2010). Unresponsive wakefulness syndrome: A new name for the vegetative state or apallic syndrome. *BMC Medicine, 8*, 68–71. doi:10.1186/1741-7015-8-68

Laureys, S., & Schiff, N. D. (2012). Coma and consciousness: Paradigms (re) framed by neuroimaging. *Neuroimage, 61*(2), 478–491. doi:10.1016/j.neuroimage.2011.12.041

Lichtensztejn, M., Macchi, P., & Lischinsky, A. (2014). Music therapy and disorders of consciousness: Providing clinical data for differential diagnosis between

vegetative state and minimally conscious state from music-centered music therapy and neuroscience perspectives. *Music Therapy Perspectives*, *32*(1), 47–55. doi:10.1093/mtp/miu001

Magee, W. L. (2005). Music therapy with patients in low-awareness states: Approaches to assessment and treatment in multidisciplinary care. *Neuropsychological Rehabilitation*, *15*(3–4), 522–536. doi:10.1080/09602010443000461

Magee, W. L. (2007). Music as a diagnostic tool in low-awareness states: Considering limbic responses. *Brain Injury*, *21*(6), 593–599.

Magee, W. L., Lenton-Smith, G., & Daveson, B. A. (2012). *Music Therapy Assessment for Awareness in Disorders of Consciousness (MATADOC): Assessment Manual and Instructions for Use.* London: Royal Hospital for Neuro-Disability

Magee, W. L., Siegert, R. J., Daveson, B. A., Lenton-Smith, G., & Taylor, S. M. (2014). Music therapy assessment tool for awareness in disorders of consciousness (MATADOC): Standardisation of the principal subscale to assess awareness in patients with disorders of consciousness. *Neuropsychological Rehabilitation*, *24*(1), 101–124. doi:10.1080/09602011.2013.844174

O'Kelly, J., James, L., Palaniappan, R., Taborin, J., Fachner, J., & Magee, W. L. (2013). *Neurophysiological and behavioral responses to music therapy in vegetative and minimally conscious states.* Switzerland: Frontiers Research Foundation. doi:10.3389/fnhum.2013.00884

O'Kelly, J., & Magee, W. L. (2013). The complementary role of music therapy in the detection of awareness in disorders of consciousness: An audit of concurrent SMART and MATADOC assessments. *Neuropsychological Rehabilitation*, (*23*)2, 287–298.

Okumura, Y., Asano, Y., Takenaka, S., Fukuyama, S., Yonezawa, S., Kasuya, Y., & Shinoda, J. (2014). Brain activation by music in patients in a vegetative or minimally conscious state following diffuse brain injury. *Brain Injury*, *28*(7), 944–950. doi:10.3109/02699052.2014.888477

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting awareness in the vegetative state. *Science*, *313*(5792), 1402.

Owen, A. M., Coleman, M. R., Menon, D. K., Johnsrude, I. S., Rodd, J. M., Davis, M. H., … Pickard, J. D. (2005). Residual auditory function in persistent vegetative state: A combined PET and fMRI study. *Neuropsychological Rehabilitation*, *15*(3–4), 290–306. doi:10.1080/09602010443000579

Perrin, F., Schnakers, C., Schabus, M., Degueldre, C., Goldman, S., Bredart, S., … Laureys, S. (2006). Brain response to one's own name in vegetative state, minimally conscious state, and locked-in syndrome. *Archives of Neurology*, *63*, 562–569.

Plum, F., and Posner, J. B. (1983). *The diagnosis of stupor and coma* (*3rd ed.*). Philadelphia: FA Davis.

Puggina, A. C. G., da Silva, M. J. P., & Santos, J. L. F. (2011). Use of music and voice stimulus on patients with disorders of consciousness. *Journal of Neuroscience Nursing*, *43*(1), E8–E16.••

R Core Team. (2014). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/.

Revelle, W. (2014). *psych: Procedures for Personality and Psychological Research.* Northwestern University, Evanston, IL.

Royal College of Physicians. (2013). *Prolonged disorders of consciousness.* National clinical guidelines. London.

Schindler v. Schiavo, 9–10 Florida Sixth Judicial Circuit. (2001).

Schnakers, C., & Majerus, S. (2012). Behavioral assessment and diagnosis of disorders of consciousness. In C. Schnakers, & S. Laureys (Eds.), *Coma and disorders of consciousness* (pp. 1–10). New York: Springer Science + Business Media.

Seel, R. T., Sherer, M., Whyte, J., Katz, D. I., Giacino, J. T., Rosenbaum, A. M., … Zasler, N. (2010). Assessment scales for disorders of consciousness: Evidence-based recommendations for clinical practice and research. *Archives of Physical Medicine and Rehabilitation*, *91*(12), 1795–1813. doi:10.1016/j.apmr.2010.07.218

Verger, J., Ruiz, S., Tillmann, B., Ben Romdhane, M., De Quelen, M., Castro, M., … Perrin, F. (2014). Beneficial effect of preferred music on cognitive functions in minimally conscious state patients. *Revue Neurologique, 170*(11), 693–699. doi:10.1016/j.neurol.2014.06.005

Verville, V. C., Sela, L., Plotkin, A., Chatelle, C., Sobel, N., & Laureys, S. (2012). Detecting signs of consciousness in severely brain injured patients with voluntary control of sniffing: A cohort study. Accepted Abstracts from the International Brain Injury Association's Ninth World Congress on Brain Injury. *Brain Injury, 26* (4–5), 721.

Wilson, S. L., Cranny, S. M., & Andrews, K. (1992). The efficacy of music for stimulation in prolonged coma: Four single-case experiments. *Clinical Rehabilitation*, *6*, 181–187.