A Systematic Review of Outcome Measures in Music Therapy

Neta Spiro[1], Giorgos Tsiris [2] Charlotte Cripps[1]

1. Nordoff Robbins Music Therapy, London

2. Nordoff Robbins Scotland

ABSTRACT

Assessment in its different guises and forms, has been a core consideration of music therapy work since the early stages of the discipline. Practitioners have devised assessment tools for different purposes such as outcome measurement. Although an increased number of tools have been published, many do not seem to be used widely in either practice or research. This situation might be connected to limited accessibility to such tools and/or the lack of centralized sources of information about them. This systematic review focuses on 26 music-therapy-specific outcome measures and identities trends and gaps in their characteristics. The results show that most measures concern work with people with autism, developmental and learning disabilities, and special needs. Most measures have been designed using pilot studies, and most original sources introducing the measures have been published since 2000 and in journal publications. Validation processes are reported in less than half of the original sources. Most measures employ observational ratings and checklists as their data collection method and have five main categories of focus: musical engagement, functioning, communication and/ or interaction, relationship, and music therapy process.

Keywords: *outcome measures, music therapy, systematic review, assessment*

Introduction

Though assessment has been integral to music therapy practice since the early 1960s (Wheeler, 2013), the drive to use outcome measures in music therapy research and practice has particularly increased in recent years. This has been encouraged by various factors, including the urge for evidence-based practice, funding expectations, as well as the belief that the use of such measures and tools can contribute to understanding about the effects and effectiveness of interventions (for a recent discussion of assessment tools, see Lipe, 2015). A greater number of studies (e.g., Aigen, 2008a, 2008b) explore whether and how music therapy works within different settings and with different client groups. Reflecting the diversity of the field in terms of theoretical approaches and practice models (Spiro, Tsiris, & Pavlicevic, 2014; Trondalen & Bonde, 2012), these studies vary not only in

terms of their methods, but also in terms of their overall focus, function, and methodological framework. Central elements of assessment-oriented studies are the selection, understanding, and use of outcome measures.

Outcome measures are relevant for assessing or comparing a person's functioning, symptoms, or presenting features when they participate in a treatment or therapy.[1] A measure can be relevant at several periods during the process of treatment or therapy:

- at, or before, the start of therapy (initial assessment);
- during therapy (alongside the process of therapy);
- at, or after, the end of therapy (often referred to as outcome assessment).

Measurements during the latter two time periods are often used to detect change in comparison with the initial assessment. Measures are often referred to as "assessment measures" (e.g. Miller, 2014) or "outcome measures" (e.g., Mackeith, Burns, & Lindeck, 2011), and the same measures may be used for both purposes. A fuller description of outcome measurement concepts and process can be found in Lipe (2015) and Spiro, Tsiris, and Cripps (in preparation), while Miller (2014) offers a description of assessment and outcomes within the wider context of arts therapies. The terms "assessment" and "evaluation" are often used interchangeably in the description of outcome measurement, with "evaluation" often referring to assessment during or after therapy. Some authors stress the differences between the two terms, and their functions have been highlighted (see Bruscia, 2005; Tsiris, Pavlicevic, & Farrant, 2014; Wheeler, 2013), but there is no agreement about these differences. For clarity, we refer to assessment throughout this study while referring to evaluation within the context of service evaluation elsewhere (Spiro & Tsiris, 2016; Spiro, Tsiris, & Pavlicevic, 2017).

Some outcome measures developed for other related professions have been used in music therapy. The Cohen-Mansfield Agitation Inventory (CMAI), for example, is often used in studies on music therapy and dementia care (see examples in Vink, Bruinsma, & Scholten [2003] and van der Steen et al. [2017]). Such measures are useful, as they speak to an audience beyond music therapists; they can encourage interdisciplinary work and may t well with the extra-musical processes and outcomes in music therapy. However, since they are not developed specifically for music therapy, such measures may be at odds with indigenous theoretical frames in the field (Aigen, 2005) and may not assess aspects uniquely relevant to its practices. This situation has led to ongoing development of discipline-specific tools. An example of a recently published outcome measure is the Music Therapy Assessment Tool for Advanced Huntington's Disease (MATA-HD)

[1] Other processes can contribute to assessment or outcome measurement (such as informal interviews), but we focus only on measures here.

(O'Kelly & Bodak, 2016), which measures patient responses to music therapy interventions across psychological, physical, social, and communication domains of functioning. Identifying the most suitable outcome measure for a given music therapy client group or setting can, in many situations, be a difficult process, and there are many possible reasons why music therapists may not use outcomes tools. Some examples of obstacles and considerations include: the format of data collection; the perceived mismatch between the nature of the tasks for assessment and the focus of the work (e.g., an emphasis on behavioral tasks that do not seem to be related to the musical relationship between clients and therapist) (Loewy, 2000); the relationship of the tool to each therapist's philosophical perspectives, client group, and work site requirements (Isenberg-Grzeda, 1988); and many professionals' lack of familiarity with the language and methods that are integral to outcome measures. There are of course many debates about whether, how, and when outcome measures are appropriate, relevant, or useful (e.g., DeNora, 2006; DeNora & Ansdell, 2014; Wigram & Gold, 2012). Engaging with these debates—while acknowledging their valuable contribution in cultivating a critical stance—is beyond the scope of this paper. The starting point here is the understanding of such measures on their own terms without judging how well regarded or useful they might be and for whom. Moreover, searching for and staying up to date with published outcome measures can be a time-consuming, daunting, and expensive task. The resulting limited awareness of existing measures can often be the initial and main difficulty in identifying and using them.

In an attempt to address some of these hurdles, a number of initiatives have emerged over the years. The newly formed International Music Therapy Assessment Consortium (Jacobsen et al., 2016) is such an example, where through cross-organizational collaboration, music therapists aim to promote the development of outcomes tools and their use within and beyond the music therapy field. Other examples include publications (e.g., Lipe, 2015) that provide listings and overviews of existing tools. Building on, and contributing to such work, we developed the Outcome Measures in Music Therapy resource (Cripps, Tsiris, & Spiro, 2016), which provides an overview of music-therapy-specific outcome measures and their characteristics. Resonating with existing resources that are not specific to music therapy, such as the Mental Health Outcomes Compendium (National Institute for Mental Health England, 2008), this resource provides a collation of core information about outcome measures. With the aim of increasing awareness and support for practitioners and researchers in the use of outcome measures, this resource is freely available online at www.nordoff-robbins.org.uk.

Aim

This study aims to identify trends and gaps in the characteristics of published music-therapy-specific outcome measures. Drawing from a subset of the measures listed in the

resource (Cripps, Tsiris, & Spiro, 2016), this study gives an overview of the content and methods of outcome measures. This includes client groups or settings, design process and validation, publication dates and types of original sources, assessment type and function, measurement method, as well as focus and presenting features or behaviors.

Methods

We implemented a systematic review as a method that involves a "detailed and comprehensive plan and search strategy derived a priori, with the goal of reducing bias by identifying, appraising, and synthesizing all relevant studies on a particular topic" (Uman, 2011, p. 57). Within healthcare contexts, systematic reviews would usually be focused on the meta-analysis and synthesis of existing findings usually in terms of effectiveness or efficacy (e.g., Carr & Wigram, 2009; Pothoulaki, MacDonald, & Flowers, 2005). Given the scope of this study, instead of exploring effectiveness and efficacy, we focus on systematic search for the outcome measures available and an analysis of their characteristics. As such, the goal was not to discuss measurement-related topics, such as reliability and validity (for details, see for example Lipe [2015] and Spiro, Tsiris, & Cripps [in preparation]), or to assess quality or applicability of outcome measures for particular contexts or uses. By understanding the current state of affairs, this study offers an opportunity to reflect on the extent to which the picture of existing music-therapy specific outcome measures relates to broader research activities within the field as well as to contemporary practice initiatives and emerging theoretical perspectives.

Data Collection

Since this study relates to its accompanying resource (Cripps, Tsiris, & Spiro, 2016), there are many overlaps between their methods. However, some differences (e.g., inclusion criteria) did occur given the different scope of the resource and this systematic review, as explained below.

Information was collected over a two-month period (October and November 2015) through online and hand searches of literature. Online searches were carried out using the following search terms in Google (including Google Scholar and Google Books) and the online library of City University, London: "music therapy measure," "music therapy assessment," "music therapy assessment measure," "music therapy outcome," "music therapy outcome measure," "music therapy rating," "music therapy scale," "music therapy rating scale." Hand searches were carried out using the literature available in the library at the Nordoff Robbins London Centre, UK. These searches included searching the references lists of books that appeared to be of relevance (i.e., Brooke, 2006; Kirkland, 2013; Lipe, 2015; Nordoff & Robbins, 2007; Oldfield, 2006; Snow & D'Amico, 2009; Wigram, Pedersen, & Bonde, 2002; Wosch & Wigram, 2007).

Inclusion Criteria

The key inclusion criterion was terminology, and no judgment about the quality or appropriateness of measures or the information reported in relation to them was made. More specifically, measures were included for further inspection if the terms "music therapy" and "measure" appeared in:

• the title of the measure, and/or

• the title or abstract of a paper describing the measure (which may or may not be the paper that introduces the measure by its authors, i.e., the original source).

If none of the above applied, the measure was included only if the original source clarified that the measure's purpose was for music therapy.

Exclusion Criteria

Initially we were interested in outcome-oriented tools (i.e., measures that assess the client in order to, for example, monitor progress, test for change or "effect" of music therapy), but during our search several measures emerged that had more than one purpose. In particular, some measures were designed for assessing client needs (needs assessment) and some looked at both outcome and needs assessment. Although the resource (Cripps, Tsiris, & Spiro, 2016)—given its more expanded remit—includes all these measures, this study focuses on a subset of these measures. More particularly, any measures that have no outcome measurement component or have been created solely to assess eligibility for music therapy are not included in the study.

The original source (i.e., the original paper that introduces the measure by its authors) was used as the primary source of information for each measure. In cases where information was inadequate, we used additional sources. On average, two sources (including the original source where possible) were used for information for each measure. The maximum number of sources used for the description of a measure was six. Any additional sources that were identified along the way were simply cited in the resource (Cripps, Tsiris, & Spiro, 2016) and not used for the description of the measures. This range of sources reflects the range of detail available for each measure. Measures were also excluded if at least one of the following applied:

- Access: There was not access to sufficient levels of information.

- Language: Measures where the only source that we had access to was not available in English.

- Publication status: Information found only in unpublished work. At least one publication had to exist for the outcome measure to be included.

- Discipline: The focus was not music therapy.

- Process: Despite the term “assessment” being used, there was no music therapy assessment process involved (e.g., service provider checklists that did not relate to a specific music therapy activity).

- Purpose: The word “measure” was used to refer to something different to assessment (e.g., service evaluation).

Data Analysis

Following the search for outcome measures and their original sources, key information was extracted about each measure. Where relevant (e.g., for the identification of presenting features or behaviors), a modest thematic analysis took place to extract and summarize key information from the sources while staying close to the wording of these sources as far as possible. Ambiguities that arose during the analytic process were discussed individually with all members of the research team, and where needed we revisited the original sources for clarification. We performed spot checks across the dataset, and we did not identify any inconsistencies. A full inter-rater reliability check was not implemented. Although this might be seen as one of the study’s weaknesses as discussed later, this check was not deemed necessary given the relatively straightforward nature of the data analysis protocol.

The kinds of information extracted for each measure and their description and coding are shown in Table 1, which draws heavily from the resource (Cripps, Tsiris, & Spiro, 2016). These are presented alongside their respective results area in order to facilitate the accessibility of the results.

Conventions about how to describe such measures are varied, and the terminology used is inconsistent across the sources. We therefore use the descriptions by the measures’ authors unless we do not have access to the original source, in which case we use the wording appearing in other sources. In certain cases, we have changed some wording that would perhaps not be considered appropriate today (e.g., “mentally retarded patients”; see Wasserman, Plutchik, Deutsch, & Taketomo, 1973). To facilitate the link between this study and the language used in its accompanying UK-based online resource (Cripps, Tsiris, & Spiro, 2016), we use the same terminology in both. Therefore, we use terms such as “autism,” “learning disability,” and “dementia” instead of the US equivalent or updated diagnostic terms (e.g., “autistic spectrum disorders” [ASD] instead of “autism,” or “neurocognitive disorders” [NCD] instead of “dementia”).

Results

Twenty-six outcome measures were identified based on the study’s inclusion and exclusion criteria. These measures were grouped into eight categories and client groups according to the population or setting for which they were developed (Table 2). No predefined categories were imposed. Instead, we assigned each measure to one category

following the information in the sources and our own knowledge and experience of some of the measures. Therefore, measures are listed only once, though they may be applicable in different categories. This results in a conservative representation of the client groups for which each measure may be relevant. For example, the Improvisational Assessment Pro les (IAPs) (Bruscia, 1987) were originally developed for people with learning and severe emotional difficulties. Since then, they have been used in studies of parental competencies for children in need of care (Jacobsen & Wigram, 2007). Similarly, the Nordoff-Robbins Scales (Nordoff & Robbins, 1977) were originally developed primarily for children with autism and have since been used with adolescent boys diagnosed with a behavioral disorder and/or emotional disorder (McIntyre, 2007).

In reporting the findings about these outcome measures, we cross-refer to examples of measures. These are named using their abbreviation (instead of reference) to allow cross-checking with Table 2. A situated summary of the findings in relation to other initiatives in the field is provided in the discussion.

Demographics

Some of the measures were developed for particular categories or client groups, such as "clients with disorders of consciousness" (MATADOC), and others were more broadly conceived, such as "children undergoing music therapy" (Music Therapy Star).

Though some measures are relevant to more than one category, as explained above, their distribution between different categories or client groups is nevertheless striking (Figure 1, see also Table 2). Over half of the outcome measures concern two closely related categories: autism, developmental and learning disabilities (n = 11, 42%), and special needs (n = 3, 11.5%). Two other relatively large categories are mental health (n = 5, 19%), and geriatric and dementia (n = 3, 12%). Within the mental health category, the majority of measures focus on client groups experiencing "emotional disturbances" (3 out of 5). All remaining categories have only one outcome measure each.

Half of the measures were not developed for a particular setting. Nevertheless, a number of settings is reported sometimes (e.g., school, psychiatric, and hospital settings), and this seems to relate to each measure's category/client group (see Table 2). Over two-thirds of the measures (n = 18, 69%) were developed for children (Figure 2). Four of these measures are also applicable for adolescents and/or adults. In addition to those measures that explicitly refer to children, the category "children" includes the IAPs that refer to the developmental age of clients (18 months minimum) and the IMCAP-ND, which was informed by measures developed for children. In the case of 13 Areas of Inquiry, adults are involved in their capacity as parents.

Design Process

The varied processes of designing the measures included: pilots (n = 7, 27%) and other

research studies (n = 10, 38%), clinical work (n = 5, 19%), and literature reviews (n = 1, 4%). In all cases, descriptions varied dramatically. Examples of scales developed as part of research studies include the MiDAS, which included focus groups and interviews as well as expert and peer consultations "to maximize its content validity" (McDermott, Orgeta, Ridder, & Orrell, 2014, p. 232). The scale was then used by music therapists and care home staff who completed weekly MiDAS ratings to allow for testing of reliability and validity (McDermott, Orgeta, Ridder, & Orrell, 2014).

Two measures (MTCS and IMCAP-ND) stated that their development was informed by a particular theoretical framework. For example, the MTCS was informed by a psychodynamic theoretical framework.

A few scales were informed by a particular music therapy approach (e.g., the Nordoff-Robbins Scales I-III were developed as part of the Nordoff-Robbins approach, while Neurologic Music Therapy informed the development of MATADOC).[2] Eight measures were informed by other scales, including some music-therapy-specific ones. The Music Therapy Checklist, for example, is derived from a selection of behaviors in the Music Therapy Coding Scheme. On the other hand, the Nordoff-Robbins Scales I-III were influenced by the scales developed by Ruttenberg et al. (1966). The latter non-music therapy scales document behavioral changes in children with autism, including assessment of stages of autonomous behavior.

Table 1

Kinds of Information, Description and Coding, and Associated Results Sections

Information	Description and coding	Results section
Client group	As specified in the sources used. Where possible and relevant, we also state which client group was used for piloting or testing a measure.	Demographics
Age group	As specified in the sources used. Where possible and relevant, we also state which age group was used for piloting or testing a measure.	
Setting	When explicitly stated or specified in the sources, the setting where a measure was piloted is included. We code "not specific" when authors describe the measure to be useful for a variety of settings.	
Design		

[2] A further outcome measure (GIMR) (Bruscia, 2000) that is based on the Guided Imagery and Music approach to music therapy was not included in the analysis because we did not have access to sufficient detail.

process	Methods used to develop the measure. These include steps such as conducting surveys (e.g., Langan, 2009), being informed by previous scales (e.g., Nordoff & Robbins, 1977), running focus groups (e.g., McDermott, Orgeta, Ridder, & Orrell, 2014), and doing pilot projects (e.g., Jacobsen & McKinney, 2015). We also make a note of cases in which it seems that particular approaches to music therapy had informed the design process.	Design process Validation
Validation (original source)	Whether the original source introducing the tool does so with the claim that it has been validated. Usually this implies that the psychometric properties of the measure have been investigated:We code "yes" if the psychometric properties have been reported on as favorable.	
Validation method (original source)	We code "no" if the authors have explicitly reported that validation has not been carried out.We code "inconclusive" if investigations into the measure's psychometric properties were inconclusive.	
Validation (sources following original)	We code "not speci ed" if no such investigations were reported in the source. Reporting on formal validation, e.g., test re-test reliability (e.g., Douglass, 2006), as described by the authors of the original source.[3] References to sources following the original one, which report on formal validation methods.	
Publication date	Publication date of the original source.	Publication dates and types of original sources
Publication type	Publication type of the original source, such as journal article or research thesis.	
Assessment type	This study includes only measures that assess outcome ("outcome assessment"), such as measures to monitor progress, test for change or "effect" of music therapy. In some cases, authors primarily suggest that a measure is relevant for "needs assessment" (as a screening and usually prior to formal music therapy to inform next steps of music therapeutic intervention)	Assessment type and function

[3] For discussion of different aspects of validity and reliability and their uses in music therapy, see Spiro, Tsiris, and Cripps (in preparation).

	but also comment about monitoring change over time. Likewise, some authors primarily describe a measure as relevant to outcome assessment and also comment that it can be useful for treatment planning. In such cases, the assessment type of these measures is labeled as "outcome/needs assessment." In all cases, we interpret the information provided in the sources and assign the labels we see as most appropriate.	
Measurement method	Method of collecting information such as behavioral checklists, rating scales, and grids.	Measurement method
Focus and presenting features or behaviors	Characteristics that the measures focus on. For example: attention; length of playing; interaction; mobility; rhythmic synchrony; following changes; boundary; engagement (Snow, 2009).We look at the stated purpose and presenting features/behaviors of each outcome measure as summarized in the resource (Cripps, Tsiris, & Spiro, 2016). The focus in terms of features/behaviors in the AQR-instrument, for example, is on "relationship," and its purpose is summarized as: "[...] to assess the quality of relationships and thus helps to evaluate the work of music therapy" (Schumacher & Calvet-Kruppa, 1999, p. 188).	Focus and presenting features or behaviors

Table 2

Category and Client Group, Outcome Measure and Original Source

Category and Client Group	**Outcome Measure**	**Original Source**
Autism, developmental and learning disabilities	Assessment of the Quality of Relationship Instrument (AQR-instrument)	Schumacher & Calvet-Kruppa (1999)
	Improvisational Assessment Pro les (IAPs)	Bruscia (1987)
	Individual Music-Centered Assessment Pro le for Neurodevelopmental Disorders (IMCAP-ND)	Carpente (2013)

	Music Therapy Assessment Tool for Adults with Developmental Disabilities (Music Therapy Assessment Tool)	Snow (2009)
	Music Therapy Checklist	Raglio, Traficante, & Oasi (2007)
	Music Therapy Coding Scheme (MTCS)	Raglio, Traficante, & Oasi (2006)
	Music Therapy Evaluation Scale	Wasserman, Plutchik, Deutsch, & Taketomo (1973)
	Nordoff-Robbins Scale I: Child-Therapist(s) Relationship in Coactive Musical Experience	Nordoff & Robbins (1977)
	Nordoff-Robbins Scale II: Musical Communicativeness Nordoff-Robbins Scale III: Musicing: Forms of Activity, Stages and Qualities of Engagement	Nordoff & Robbins (1977) Nordoff & Robbins (1977)
	13 Categories of Response	Nordoff & Robbins (1971)
Children receiving music therapy	The Music Therapy Star	Mackeith, Burns, & Lindeck (2011)
Child protection: families at risk	Assessment of Parenting Competencies–Revised (APC-R)	Jacobsen & McKinney (2015)
Disorders of consciousness	Music Therapy Assessment Tool for Awareness in Disorders of Consciousness (MATADOC)	Magee (2007)
Geriatric and dementia	Geriatric Music Therapy Clinical Assessment	Hintz (2000)
	Music-Based Evaluation of Cognitive Functioning (MBECF)	Lipe (1994)
	Music in Dementia Assessment Scales (MiDAS)	McDermott, Orgeta, Ridder, & Orrell (2014)
Hospital Mental health	Pediatric Inpatient Music Therapy Assessment Form (PIMTAF)	Douglass (2006)

	Beech Brook Music Therapy Assessment	Layman, Hussey, & Laing (2002)
	Music Interaction Rating Scale (MIR(S))	Pavlicevic (1991)
	Music Therapy Assessment for Emotionally Disturbed Children	Goodman (1989)
	Music Therapy Rating Scale (MAKS)	von Moreau (1996)
	13 Areas of Inquiry	Loewy (2000)
Special needs	Individualized Music Therapy Assessment Profile (IMTAP)	Baxter et al. (2007)
	Music Therapy Communication and Social Interaction Scale–Group (MTCSI)	Guerrero et al. (2014)
	Music Therapy Special Education Assessment Tool	Langan (2009)

Validation

For 38% (n = 10) of the measures, a validation process was reported in the original sources. In the remaining cases, the authors explicitly report that validation has not been carried out (n = 6, 23%) or information regarding validation was not given (n = 8, 31%). Though the earliest measures were published in the 1970s, those that do report validation have only been published since 2005.

The validation method in 90% (9 out of 10) of the validated measures included testing for inter-rater reliability. The raters involved varied in terms of their professional background (e.g., for PIMTAF raters were music therapists) and/or training in the use of the measures (e.g., raters for the MIR(S) were trained). In many cases, however, such information is not clearly provided.

For 31% of all the measures (8 out of 26), a follow-up validation procedure was reported in a source following the original one. In 62% of these cases (5 out of 8), no validation was reported in the original source, whereas in 75% (6 out of 8) at least one of the authors of the original source is also an author on the follow-up publication. The time span between the original source and the follow-up publication reporting on validation differs dramatically, from 1 year (MiDAS, later validated by McDermott, Orrell, & Ridder, 2015) to 33 years (Nordoff-Robbins Scale I, later validated by Mahoney, 2010).

Publication Dates and Types of Original Sources

Over half of the original sources were published since 2000 (n = 15, 58%) and in the form of journal articles (n = 14, 54%). Several outcome measures were developed as part of a research thesis (Figure 3). For example, the MiDAS was originally developed as part of McDermott's (2014) PhD thesis, and the APC-R was a development of the original Assessment of Parenting Competencies (APC), which was developed as part of Jacobsen's (2012) PhD thesis. In three cases, such theses form the original source consulted for this study. In terms of original sources reporting on validation, 70% (7 out of 10) were published in journals and 70% were published since 2000.

Assessment Type and Function

Most measures (n = 19, 73%) focus on outcome assessment, and the rest on both needs and outcome assessment. In 46% (n = 12) of the measures, no function additional to assessment was featured. In the remaining measures (n = 14), two main categories of function were represented and could be relevant to the same measures: i) clinical work and treatment planning (n = 10, 71%), and ii) screening and diagnostic assessment (n = 6, 43%). Different levels of detail were given about the function of each measure, and as such multiple interpretations of the material are possible. In some cases, the function might be implicit to the type of assessment of each measure (i.e., outcome or needs assessment). For the purposes of this study, we concentrated on the information summarized in the "purpose" sections of the resource only. Although the assessment function of measures is explicitly mentioned in the purpose of certain measures, we assumed that this applies in all cases given their default function. In some cases, the assessment elements of the measures were related to particular features of the setting. The Music Therapy Special Education Assessment Tool, for example, assesses the music therapeutic process and progress in relation to special education settings and curriculum.

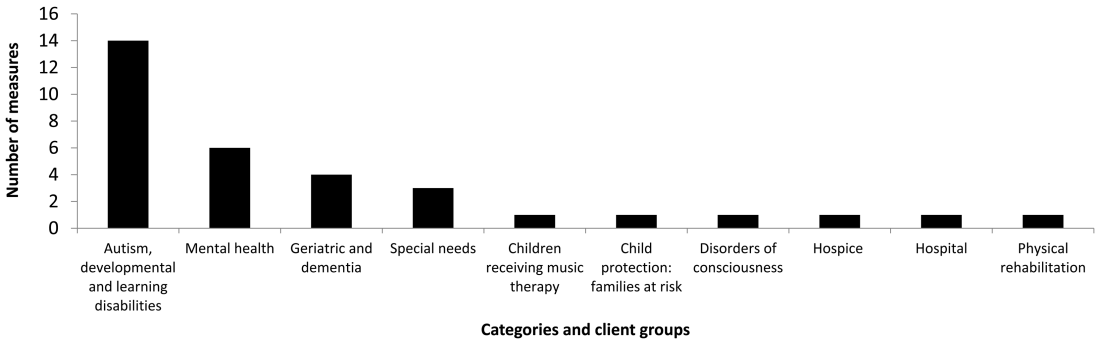

Figure 1. Outcome measures per categories and client groups.

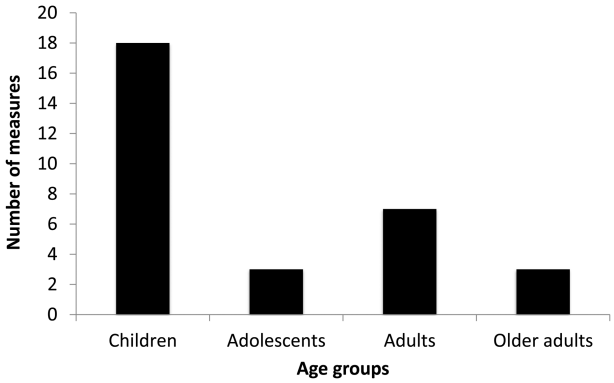

Figure 2. Outcome measures per age groups.

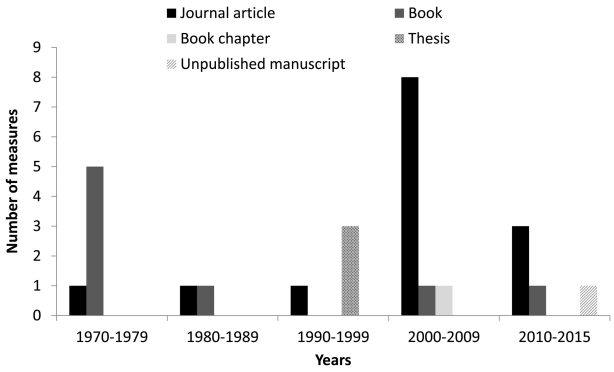

Figure 3. Outcome measures per publication date and type of original source.

Measurement Method

Observational ratings and checklists are by far the most common data collection method used in the measures (n = 22, 85%). Methods of recording observations are wide ranging: some measures use checklists (e.g., Music Therapy Checklist), sometimes with open-ended comments (e.g., Music Therapy Assessment for Emotionally Disturbed Children), others include both rating scales and qualitative ratings (e.g., Music Therapy Assessment Tool), while others have grids (e.g., Nordoff-Robbins Rating Scales I-III) or spatial representations to fill in (e.g., Music Therapy Star).

Often observation relates to free improvisation during music therapy sessions (e.g., IAPs, Nordoff-Robbins Rating Scales I-III, and MIR(S)). Twenty-three percent (n = 5) of the measures that collect data via observational ratings and checklists report use of video footage. Some measures are rated or coded after sessions as a means of analysis, using video footage of sessions (e.g., the MTCSI, the MIR(S), and the MTCS), while others are rated by hand immediately after therapy sessions. For instance, the MiDAS includes forms to be completed at different times (e.g., before and/or after) by different care home staff and music therapists (McDermott, Orrell, & Ridder, 2015).

In some cases, observational ratings and checklists are combined with other methods such as microanalysis (e.g., the AQR-instrument). Other non-observational methods include task-based ones (12%, n = 3) and tend to have a specific protocol to administer (e.g., Music Therapy Assessment for Disturbed Adolescents). Only one measure (13 Areas of Inquiry) uses narrative descriptions as its method of documenting information.

Focus and Presenting Features or Behaviors

Five categories (which overlapped to some extent) were identified in terms of the measures' foci: musical engagement (n = 10, 38%), functioning (n = 9, 35%), communication and/or interaction (n = 6, 23%), relationship (n = 3, 12%), and music therapy process (n = 2, 8%). Musical engagement—the most common area of focus—includes, for example, musical communicativeness, experience, and behaviors. Functioning—the second most common area of focus—includes, for example, clients' strengths, abilities, and needs, as well as their cognitive and physical functioning.

Only a few of the original sources (e.g., Carpente, 2013) clarify the duration and/or frequency of use of outcome measures, including information as to whether a measure is designed for short- or long-term assessment. The Music Therapy Checklist, for example, is described as applicable both within single sessions and for the entire course of treatment, and the IMCAP-ND is used as a pre- and post-test measure to assess client progress, on either a short- or long-term basis. In some cases, however, duration is indicated in general terms.

The categories of focus mentioned above are addressed through measurement of targeted presenting features or behaviors that may include client symptoms. Six main areas of presenting features or behaviors were identified (see Table 3). Most measures (n = 22, 85%) focus on more than one presenting feature or behavior. On average, three areas of presenting features or behaviors were identified per outcome measure, with "musical skills or participation" (n = 24, 92%) and "communication and/or interaction" (n = 18, 69%) being the most common areas.

Moreover, focus tends to be on behaviors that take place during music-making, which are in turn considered to have implications for clinically relevant objectives, such as levels of interaction and communicativeness (e.g., the Nordoff-Robbins Scales I-III; the MTCS, and the MAKS), or quality of relationship between client and therapist (e.g., the IAPs and the AQR instrument). Indeed, behaviors observed are often concerned with musical responsiveness, such as perceived rhythmic synchrony (Music Therapy Assessment Tool), musical attention (IMCAP-ND), non-verbal communication skills (APC-R), as well as play and creativity (Music Therapy Star). Other measures use music listening, verbal, singing, and rhythm tasks to assess cognitive functioning in older adults with dementia (e.g., MBECF).

Table 3

Areas of Presenting Features or Behaviors

Areas of presenting features or behaviors	**Examples per area of presenting features or behaviors**
Musical skills or participation (n = 24)	Length of playing and rhythmic synchrony (Music Therapy Assessment Tool), sonorous musical communication (MTCS), and qualities of participation and of restiveness (Nordoff-Robbins Scale I).
Communication and/or interaction (n = 18)	Nonverbal and verbal communication (e.g., Music Therapy Checklist), and responses to singing (13 Categories of Response).
Cognitive (n = 9)	Attention (Music Therapy Star).
Physical (n = 7)	Body movement (e.g., Nordoff-Robbins Scale II), and gross and fine motor skills (e.g., IMTAP).
Social (n = 7)	Behavioral/social functioning (e.g., Beech Brook Music Therapy Assessment)
Emotional (n = 9)	Emotional expression (e.g., Music Therapy Special Education Assessment Tool) and emotional wellbeing (e.g., Music Therapy Star).
Other (n = 8)	Boundaries (e.g., Music Therapy Assessment Tool for Adults with Developmental Disabilities) and level of support needed (e.g., IMCAP-ND).

Discussion

Though music therapy outcome measures were already created in the 1970s, there has been a recent increase in the number of measures and in the proportion of which are validated as part of their development. These trends may be related to some developments in music therapy research, including an evolving focus on evidence-based practice and related methods (Edwards, 2002, 2005). This focus is also reflected in topics of research theses (for a content analysis of USA-based music therapy theses and dissertations, see Flores, 2013). Indeed, several outcome measures were developed as part of a research thesis. These trends also coincide with the establishment and development of peer-reviewed journals in music therapy (Tsiris, Spiro, & Pavlicevic, 2014).

Looking at the broader landscape of the music therapy profession and discipline, music therapists are now in a better position to, and are often expected to, use outcome measures in different contexts, as well as for different purposes and target audiences. At the same time, there seems to be an increase in demand for using validated tools in music therapy assessment and research. With a direct impact on our ways of constructing and testing such tools in order to enhance their integrity, aspects, such as accessibility and consistency of use, are increasingly key characteristics of outcome measures in terms of their development and dissemination. Validation and testing for reliability indeed refer to the processes of checking whether a tool measures what was intended and whether it is used consistently by different users (e.g., usually using tests of inter-rater reliability) (Streiner, Norman, & Cairney, 2014). It is worth mentioning here that most measures do not seem to require training for their use and do not necessarily require the user to be a qualified music therapist. MATADOC (Magee, 2007) is one of the exceptions where training on the measure in addition to music therapy qualification is required.

As many music therapy practitioners and researchers will have experienced in their daily work and communication with service providers, funders, and policymakers, however, the term "validation" is often used more broadly and often connected to a broad sense of using tools that are "recognized." This broadening of the term often comes with a less informed understanding of what validation means as well as of the validating process. This situation can create confusion in the understanding and use of outcome measures as well as in the reporting of their outcomes.

As shown in this study, the music-therapy-specific outcome measures are concentrated on a relatively small number of categories/client groups or settings, though some may be applicable to a wider range or may not be conceived as setting- specific. More specifically, measures concentrate on a subset of client groups (autism, developmental and learning disabilities [42%], and special needs [11.5%]; mental health [19%]; geriatric and dementia [12%]). In terms of client age groups, most measures were developed for children. Within client group areas, there seems to be a concentration on particular aspects (e.g., in the mental health category there is a focus on client groups experiencing "emotional disturbances"). Though most were developed with a client group or condition in mind, almost half of the measures were not developed for a particular setting, indicating that priority is given to client group/condition rather than the context of work.

This pro le of existing measures appears to align with the kinds of client groups that are reported both in surveys of the profession and of the literature in the field in the UK. According to a recent survey of the UK workforce (Carr, Tsiris, & Swijghuisen Reigersberg, 2017), for example, music therapists reported working most with clients with learning disabilities (74%), autistic spectrum (71%), and emotional and behavioral disorders (62%). Similarly, a content analysis of the British Journal of Music Therapy (Tsiris, Spiro, & Pavlicevic, 2014) showed that people with mental health problems (17%) and people with learning difficulties (15%) are the largest groups represented in

practice-based, research, and evaluation papers. In contrast, the emphasis of outcome measures on children does not align with the trend in publications in the British Journal of Music Therapy which focus on adults (20–59 years old). Although the emphasis on particular aspects within client group areas (e.g., in the mental health category there is a focus on client groups experiencing "emotional disturbances") may be surprising given that music therapists work with people with a range of psychiatric conditions such as schizophrenia and depression, this may relate to foci that lend themselves to outcome measures (such as a focus on positive change or improvement or on reduction of symptoms). Overall, given the small number of outcome measures, it is not surprising that some areas of work do not seem to be explicitly covered by them yet. Perhaps these gaps will be filled in time by the development of measures that address client age groups and aspects of music therapy work that are relatively underrepresented or absent in the current outcome measures. Such developments, however, need to consider the pro le and needs of their target audience, which may differ from the UK-related characteristics that we offer as an example here.

Most measures are not described as being relevant only to specific music therapy approaches and do not seem to be explicitly informed by particular theoretical frameworks or approaches to music therapy. Some exceptions, however, include the Nordoff-Robbins Scales (Nordoff & Robbins, 1977), which are informed by the Nordoff-Robbins approach to music therapy, and the MATADOC (Magee, 2007), which is informed by Neurologic Music Therapy (see also the GIMR [Bruscia, 2000], which is specific to Guided Imagery and Music). This limited reference to particular music therapy approaches can be seen in different ways. On one hand, outcome measures appear to contribute to the wider discipline and knowledge base without having the goal of addressing practices or concepts that are specific to certain approaches. Also, the outcomes of such measures are often used within multidisciplinary contexts where the priorities might include comparison with other interventions or communication about change in a language that is shared. Nevertheless, given the wide range of approaches within music therapy and their respective methods (Spiro, Tsiris, & Pavlicevic, 2014), this lack of approach-specific measures seems striking and can potentially give the impression that the field is more homogeneous than it is (Tsiris, 2013). The question of whether it is important to have approach-specific measures or not is, of course, relevant to a wider question regarding the usefulness of music-therapy-specific measures altogether.

In these debates, considerations regarding the role and contribution of music therapy theory to the development not only of outcome measures but also of practice and of particular techniques in music therapy seem relevant. Stige (2015) argues that there has been a practice turn in music therapy theory with implications on the future developments of the discipline and profession. He suggests that the emerging family of "'practice turn theories' in music therapy would highlight the social and performed nature of music's help, where practice is a site of knowing, not just a site for application of knowledge"

(Stige, 2015, p. 4). These considerations are situated within a wider spectrum of discussions regarding the importance of developing music therapy theory in pace with practice and research (Ghetti, 2015), as well as the role of developing indigenous music therapy theories[4] that address the salient elements of creative therapeutic practice (Aigen, 1991, 2005; Daveson, O'Callaghan, & Grocke, 2008). Such debates and considerations can help avoid schisms between research, theory, and practice. As such, they can contribute to the development of outcome measures that are useful and relevant to music therapy practitioners.

Most of the measures focus on "outcome assessment," i.e., measures that assess the client in order to, for example, monitor progress, or test for change or "effect" of music therapy. On the other hand, some of the measures were also designed for assessing client needs ("needs assessment"). Five categories were identified in terms of the measures' foci: functioning, musical engagement, relationship, communication and/or interaction, and music therapy process. These foci are traced by looking at presenting features or behaviors, with the most common presenting feature or behavior being "musical skills or participation." Within this, two perspectives dominate: From one perspective, focus tends to be on behaviors that typically take place during music-making, which are in turn considered to have implications for clinically relevant objectives, such as levels of interaction and communicativeness, or quality of relationship between client and therapist. From another perspective, measures assess frequency of behaviors symptomatic of a particular condition that occur during music-making or tasks.

This distinction between outcome and needs assessment, as well as their different foci, is crucial in deciding which measure might be appropriate in each case. In terms of focus, the concentration on musical skills or participation is not surprising given that we consider only music-therapy-specific outcome measures. At the same time, however, this finding highlights the specific gap that such discipline-specific measures can address.

This study has a number of weaknesses and limitations that could be addressed in future studies. These include the use of additional online and hand searches, as well as the expansion of the current inclusion/exclusion criteria. The latter would enable the inclusion of relevant studies that may not include the terms "music therapy" and "measure" in their title or abstract. Such studies have been omitted here, and as such the findings should not be treated as representative of all outcome measures in the field. Furthermore, future studies could include a full inter-rater check.

Looking ahead, a number of questions emerge from the findings of this study regarding

---

[4] Since its initial conception by Aigen (1991, 2005), the notion of indigenous music therapy theory has been developed to refer to "knowledge emergent from music therapy relationships which are able to be plausibly generalised across settings and time, allowing for the development of theory that is idiosyncratic to the field of music therapy" (Daveson, O'Callaghan, & Grocke, 2008, p. 280).

both the outcome measures themselves and their relationship with music therapy practices. One set of questions concerns the investigation of the use and future development of measures in practice and research. The list of outcome measures alongside their characteristics can be used as a starting point to ask such questions. For example: As mentioned earlier, some measures are used with client groups beyond those for which the measures were originally developed. However, how common this is remains an open question. Related questions include, for example, which of the measures are being used by practitioners, in what contexts, and for what purposes? Do practitioners use measures as part of their own analyses of music therapy work? Does whether they are validated measures make a difference to whether or not outcome measures are used? What other criteria affect choice—length, availability, or others? Are the measures used for communication with funders and policymakers? Other related questions include the representation of clients' perspectives and ratings (e.g., self-report) as well as the accessibility of measures by clients and their use for co-shaping the direction of their sessions with the therapist.

Similar questions emerge in terms of the selection and use of outcome measures in research—whether these are used independently or in combination with measures that are not specific to music therapy. One of the most likely types of research projects in which measures are likely to feature is the randomized controlled trial. In this context in particular, given the increase of international multicenter studies in music therapy, systematic translational processes are becoming more important. Such international studies are likely to require outcome measures not published in English and suggest the need to consider the translation and cross-cultural adaptation of measures (Ridder, McDermott, & Orrell, 2015). This is particularly relevant when measures are transferred and/or translated to contexts that are different from the ones within which they were originally developed. Indeed, this study did not consider non-English-language outcome measures. This study therefore does not allow for an international view or a comparison between measures from different geographical contexts, and this can certainly be a future research focus. In addition to the translation and transferability of outcome measures in different languages and cultures, however, it would be worth considering the extent to which outcome measures consider client's ethnicity, language, or culture as a part of the assessment process.

In closing, it is worth acknowledging that in addition to the outcome measures analyzed in this study, music therapists have been involved in the development of a number of other measures that are not necessarily music therapy specific. Two examples are the Healthy-Unhealthy Music Scale (HUMS) (Saarikallio, Gold, & McFerran, 2015) and the Interest in Music scale (IiM) (Gold, Rolvsjord, Mössler, & Stige, 2013). Both scales can be used as an outcome measure not only in music therapy, but also in other music-related interventions. HUMS, for example, can be used as an outcome measure for "therapy interventions using music that deal with the treatment of depression and/or developing healthy relationships with music" (Saarikallio, Gold, & McFerran, 2015, p. 216). In

addition to the potential interdisciplinary direction and application of modern outcome measures developed by music therapists, these scales also point toward a potentially wider focus of such measures. This wider focus appears to refer not only to immediate "benefits" of music therapy interventions, but also to their possible negative impact (e.g., HUMS) as well as their relevance to clients' everyday lives beyond the music therapy context (e.g., IiM). These developments seem to be connected to recent shifts in the field, such as the development of community music therapy and a focus on the "everyday" (Ansdell, 2014; Ansdell & Stige, 2016; Bonde, Ruud, Skånland, & Trondalen, 2013), and future research in the use of outcome measures would benefit from considering their interrelationships with these wider developments.

Acknowledgments

We express our appreciation to Camilla Farrant and other members of the Nordoff Robbins research team who helped in the development of the Outcome Measures in Music Therapy resource that formed the basis for this study. We also thank the many music therapists, conversations with whom contributed to the identification of the need for such a resource.

References

Aigen, K. (1991). The roots of music therapy: Towards an indigenous research paradigm (PhD dissertation, New York University). Retrieved from http://steinhardt.nyu.edu/scmsAdmin/uploads/005/496/AigenKen1991.pdf

Aigen, K. (2005). Music-centered music therapy. Gilsum, NH: Barcelona Publishers.

Aigen, K. (2008a). An analysis of qualitative music therapy research reports 1987–2006: Articles and book chapters. The Arts in Psychotherapy, 35(4), 251–261.

Aigen, K. (2008b). An analysis of qualitative music therapy research reports1987–2006: Doctoral studies. The Arts in Psychotherapy, 35(5), 307–319.

Ansdell, G. (2014). How music helps in music therapy and everyday life. Farnham: Ashgate.

Ansdell, G., & Stige, B. (2016). Community music therapy. In J. Edwards (Ed.), The Oxford handbook of music therapy (pp. 595–621). Oxford: Oxford University Press.

Baxter, H. T., Berghofer, J. A., MacEwan, L., Nelson, J., Peters, K., & Roberts, P. (2007). The individualized music therapy assessment pro le. London: Jessica Kingsley Publishers.

Bonde, L. O., Ruud, E., Skånland, M. S., & Trondalen, G. (Eds.). (2013). Musical life stories: Narratives on health musicking. Oslo: Centre for Music and Health, Norwegian

Academy of Music.

Brooke, S. L. (2006). Creative arts therapies manual: A guide to the history, theoretical approaches, assessment, and work with special populations of art, play, dance, music, drama, and poetry therapies. Springfield, IL: Charles C. Thomas.

Bruscia, K. (1987). Improvisational models of music therapy. Springfield, IL: Charles C. Thomas.

Bruscia, K. (1995). Client assessment in music therapy. Unpublished manuscript.

Bruscia, K. (2000). A scale for assessing responsiveness to guided imagery and music. Journal of the Association of Music and Imagery, 7, 1–7.

Bruscia, K. (2005). Research topics and questions in music therapy. In B. Wheeler (Ed.), Music therapy research (pp. 81–93). Gilsum, NH: Barcelona Publishers.

Carpente, J. (2013). The individual music-centered assessment pro le for neurodevelopmental disorders (IMCAP-ND): A clinical manual. North Baldwin, NY: Regina Publishers.

Carr, C., Tsiris, G., & Swijghuisen Reigersberg, M. (under review). Understanding the present, re-visioning the future: An initial mapping of music therapists in the United Kingdom.

Carr, C., & Wigram, T. (2009). Music therapy with children and adolescents in main-stream schools: A systematic review. British Journal of Music Therapy, 23(1), 3–18.

Cripps, C., Tsiris, G., & Spiro, N. (Eds.). (2016). Outcome measures in music therapy: A resource developed by the Nordoff Robbins research team. London: Nordoff Robbins. Retrieved from www.nordoff-robbins.org.uk

Daveson, B., O'Callaghan, C., & Grocke, D. (2008). Indigenous music therapy theory building through grounded theory research: The developing indigenous theory framework. The Arts in Psychotherapy, 35(4), 280–286.

DeNora, T. (2006). Evidence and effectiveness in music therapy: Problems, possibilities and performance in health contexts. British Journal of Music Therapy, 20(2), 81–99.

DeNora, T., & Ansdell, G. (2014). What can't music do? Psychology of Well-Being: Theory. Research and Practice, 23(4), 1–10. Retrieved from http://www.psywb.com/content/pdf/s13612-13014-10023–13616.pdf.

Douglass, E. T. (2006). The development of a music therapy assessment tool for hospitalized children. Music Therapy Perspectives, 24(2), 73–79.

Edwards, J. (2002). Using the evidence based medicine framework to support music therapy posts in healthcare settings. British Journal of Music Therapy, 16(1), 29–34.

Edwards, J. (2005). Possibilities and problems for evidence-based practice in music therapy. The Arts in Psychotherapy, 32(4), 293–301.

Flores, L. (2013). A content analysis of music therapy theses and dissertations

from 2000 to 2012 from AMTA approved graduate programs (Master's thesis, Western Michigan University). Retrieved from http://scholarworks.wmich.edu/ masters_theses/122/

Ghetti, C. M. (2015). Maintaining the dialogue of in uence: Developing music therapy theory in pace with practice and research. Approaches: Music Therapy & Special Music Education, Special Issue 7(1), 30–37. Retrieved from http:// approaches.gr/wp-content/uploads/2015/08/3-Approaches_712015_Ghetti_ Article.pdf

Glynn, N. J. (1992). The music therapy assessment tool in Alzheimer's patients. Journal of Gerontological Nursing, 18(1), 3–9.

Gold, C., Rolvsjord, R., Mössler, K., & Stige, B. (2013). Reliability and validity of a scale to measure interest in music among clients in mental health care. Psychology of Music, 41(5), 665–682.

Goodman, K. (1989). Music therapy assessment of emotionally disturbed children. The Arts in Psychotherapy, 16, 179–192.

Guerrero, N., Hummel-Rossi, B., Turry, A., Eisenberg, N., Selim, N., Birnbaum, J., Marcus, D., & Ritholz, M. (2014). Music therapy communication and social interaction scale–group.

Hintz, M. R. (2000). Geriatric music therapy clinical assessment: assessment of music skills and related behaviors. Music Therapy Perspectives, 18(1), 31–40.

Isenberg-Grzeda, C. (1988). Music therapy assessment: A re ection of professional identity. Journal of Music Therapy, 25(3), 156–169.

Jacobsen, S. (2012). Music therapy assessment and development of parental competences in families where children have experienced emotional neglect: An investigation of the reliability and validity of the tool, Assessment of Parenting Competencies (APC) (PhD thesis, Aalborg University, Denmark). Retrieved from http://vbn.aau.dk/en/publications/music-therapy-assessment-and-development-of-parental-competences-in-families-where-children-have-experienced-emotional-neglect(f3f4480f-f313-476a-b009-95c8300558e9).html

Jacobsen, S. L., Magee, W., Storm, S., Thomas, D., O'Kelly, J., Wosch, T., & AlaRuona, E. (2016). Music therapy assessment: Bridging gaps. Nordic Journal of Music Therapy, 25(sup 1), 92–92.

Jacobsen, S. L., & McKinney, C. H. (2015). A music therapy tool for assessing parent-child interaction in cases of emotional neglect. Journal of Child and Family Studies, 24, 2164–2173.

Jacobsen, S., & Wigram, T. (2007). Music therapy for the assessment of parental competencies for children in need of care. Nordic Journal of Music Therapy, 16(2), 129–143.

Kirkland, K. (Ed.). (2013). International dictionary of music therapy. New York: Routledge.

Langan, D. (2009). A music therapy assessment tool for special education: Incorporating education outcomes. Australian Journal of Music Therapy, 20, 78–98.

Layman, D. L., Hussey, D. L., & Laing, S. J. (2002). Music therapy assessment for severely emotionally disturbed children: A pilot study. Journal of Music Therapy, 39(3), 164–187.

Lipe, A. (1994). The use of music performance tasks in the assessment of cognitive functioning among older adults with dementia (PhD thesis, University of Maryland College Park).

Lipe, A. (2015). Music therapy assessment. In B. Wheeler (Ed.), Music therapy handbook (pp. 76–90). New York: Guildford Press.

Loewy, J. (2000). Music psychotherapy assessment. Music Therapy Perspectives, 18(1), 47–58.

Mackeith, J., Burns, S., & Lindeck, J. (2011). The Music Therapy Star: The outcomes star for children in music therapy. Triangle Consulting Social Enterprise, 2–18.

Magee, W. L. (2007). Development of a music therapy assessment tool for patients in low awareness states. NeuroRehabilitation, 22(4), 319–324.

Mahoney, J. F. (2010). Interrater agreement on the Nordoff-Robbins evaluation scale I: Client-therapist relationship in musical activity. Music and Medicine, 2(1), 23–28.

McDermott, O. (2014). The development and evaluation of music in dementia assessment scales (MiDAS) (PhD thesis, Aalborg University, Denmark). Retrieved from http://vbn.aau.dk/en/publications/the-development-and-evaluation-of-music-in-dementia-assessment-scales-midas(ee02eb03-d92a-4643-94b0-cb2438111515).html

McDermott, O., Orgeta, V., Ridder, H. M., & Orrell, M. (2014). A preliminary psychometric evaluation of music in dementia assessment scales (MiDAS). International Psychogeriatrics, 26(6), 1011–1019.

McDermott, O., Orrell, M., & Ridder, H. M. (2015). The development of music in dementia assessment scales (MiDAS). Nordic Journal of Music Therapy, 24(3), 232–251.

McIntyre, J. (2007). Creating order out of chaos: Music therapy with adolescent boys diagnosed with a behaviour disorder and/or emotional disorder. Music Therapy Today, VIII(1), 56–79.

Miller, C. (2014). Assessment and outcomes in the arts therapies: A person-centred approach. London: Jessica Kingsley Publishers.

National Institute for Mental Health England (2008). Mental health outcomes compendium. Retrieved from http://webarchive.nationalarchives.gov.uk/20130107105354/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_093677.pdf

Nordoff, P., & Robbins, C. (1971). Therapy in music for handicapped children. London: Gollancz.

Nordoff, P., & Robbins, C. (1977). Creative music therapy. New York: John Day.

Nordoff, P., & Robbins, C. (2007). Creative music therapy: A guide to fostering clinical musicianship (2nd ed.). Gilsum, NH: Barcelona Publishers.

O'Kelly, J. (2016). Music therapy assessment tools in practice: Challenges and opportunities. Nordic Journal of Music Therapy, 25(sup1), 143.

O'Kelly, J., & Bodak, R. (2016). Development of the music therapy assessment tool for advanced Huntington's disease: A pilot validation study. Journal of Music Therapy, 53(3), 232–256.

Oldfield, A. (2006). Interactive Music Therapy in Child and Family Psychiatry: Clinical Practice, Research, and Teaching. London: Jessica Kingsley Publishers.

Pavlicevic, M. (1991). Music in communication: Improvisation in music therapy (PhD thesis, University of Edinburgh).

Pothoulaki, M., MacDonald, R., & Flowers, P. (2005). Music interventions in oncology settings: A systematic literature review. British Journal of Music Therapy, 19(2), 75–83.

Raglio, A., Tra cante, D., & Oasi, O. (2006). A coding scheme for the evaluation of the relationship in music therapy sessions. Psychological Reports, 99(1), 85–90.

Raglio, A., Tra cante, D., & Oasi, O. (2007). Comparison of the music therapy coding scheme with the music therapy checklist. Psychological Reports, 101, 875–80.

Ridder, H. M., McDermott, O., & Orrell, M. (2015). Translation and adaptation procedures for music therapy outcome instruments. Nordic Journal of Music Therapy, 26(1) 62–78.

Ruttenberg, B. A., Dratman, M. L., Fraknoi, J., & Wenar, C. (1966). An instrument for evaluating autistic children. Journal of the American Academy of Child Psychiatry, 5(3), 453–478.

Saarikallio, S., Gold, C., & McFerran, K. (2015). Development and validation of the healthy - unhealthy music scale. Child and Adolescent Mental Health, 20(4), 210-217.

Schumacher, K., & Calvet-Kruppa, C. (1999). The "AQR"—an analysis system to evaluate the quality of relationship during music therapy: Evaluation of interpersonal relationships through the use of instruments in music therapy with profoundly developmentally delayed patients. Nordic Journal of Music Therapy, 8(2), 188–191.

Snow, S. (2009). The development of a music therapy assessment tool: A pilot study. In S. Snow & M. D'Amico (Eds.), Assessment in the creative arts therapies: Designing and adapting assessment tools for adults with developmental disabilities (pp. 47–98). Springfield, IL: Charles C. Thomas.

Snow, S., & D'Amico, M. (Eds.). (2009). Assessment in the creative arts therapies: Designing and adapting assessment tools for adults with developmental disabilities. Springfield, IL: Charles C. Thomas.

Spiro, N., & Tsiris, G. (2016). Assessment and evaluation in music therapy: Is there a difference? Nordic Journal of Music Therapy, 25(sup1), 70–71.

Spiro, N., Tsiris, G., & Cripps, C. (in preparation). "Sounds good... but, what is it?" An introduction to outcome measures with examples from music therapy.

Spiro, N., Tsiris, G., & Pavlicevic, M. (2014). Music therapy models. In W. F. Thompson (Ed.), Music in the social and behavioral sciences: An encyclopedia (pp. 772–774). Thousand Oaks, CA: Sage.

Spiro, N., Tsiris, G., & Pavlicevic, M. (2017). Re-positioning music therapy service evaluation: A case of ve Nordoff-Robbins music therapy service evaluations in neuro-rehabilitation. Nordic Journal of Music Therapy, 1–25.

Stige, B. (2015). The practice turn in music therapy theory. Music Therapy Perspectives, 33(1), 3–11.

Streiner, D. L., Norman, G. R., & Cairney, J. (2014). Health measurement scales: A practical guide to their development and use. Oxford: Oxford University Press.

Trondalen, G., & Bonde, L. O. (2012). Music therapy: Models and interventions. In R. MacDonald, G. Kreutz, & L. Mitchell (Eds.), Music, health, and wellbeing (pp. 40–62). Oxford: Oxford University Press.

Tsiris, G. (2013). Voices from the "ghetto": Music therapy perspectives on disability and music (A response to Joseph Straus's book "Extraordinary measures:

Disability in music"). International Journal of Community Music, 6(3), 333–343.

Tsiris, G., Pavlicevic, M., & Farrant, C. (2014). A guide to evaluation for arts therapists and arts & health practitioners. London: Jessica Kingsley Publishers.

Tsiris, G., Spiro, N., & Pavlicevic, M. (2014). What does the past tell us? A content analysis of the first quarter-century of the British Journal of Music Therapy. British Journal of Music Therapy, 28(1), 2–24.

Uman, L. S. (2011). Systematic reviews and meta-analyses. Journal of the Canadian Academy of Child and Adolescent Psychiatry, 20(1), 57–59.

van der Steen, J. T., van Soest-Poortvliet, M. C., van der Wouden, J. C., Bruinsma, M. S., Scholten, R. J., & Vink, A. C. (2017). Music-based therapeutic interventions for people with dementia. Cochrane Database of Systematic Reviews (5. Art. No.: CD003477). http://doi. org/10.1002/14651858.CD003477.pub3

Vink, A. C., Bruinsma, M. S., & Scholten, R. J. P. M. (2003). Music therapy for people with dementia. Cochrane Database of Systematic Reviews, Issue 4. Art. No.: CD003477. DOI: 10.1002/14651858.CD003477.pub2.

von Moreau, D. (1996). Entwicklung und Evaluation eines Beschreibungs-systems (MAKS)

zumAusdrucks- und Kommunikationsverhalten in der Musiktherapie [Development and evaluation of a rating system (MAKS) on expression and social behaviour in music therapy] (Unpublished thesis, University of Würzburg, Germany).

Wasserman, N., Plutchik, R., Deutsch, R., & Taketomo, Y. (1973). A music therapy evaluation scale and its clinical application to mentally retarded adult patients. Journal of Music Therapy, 10(2), 64–77.

Wheeler, B. (2013). Music therapy assessment. In R. F. Cruz & B. Feder (Eds.), Feders' the art and science of evaluation in the arts therapies (2nd ed.) (pp. 344–382). Springfield, IL: Charles C. Thomas.

Wigram, T., & Gold, C. (2012). The religion of evidence-based practice: Helpful or harmful to health and wellbeing? In R. MacDonald, G. Kreutz & L. Mitchell (Eds.), Music, health, and wellbeing (pp. 164–182). Oxford: Oxford University Press.

Wigram, T., Pedersen, I. N., & Bonde, L. O. (2002). A comprehensive guide to music therapy: theory, clinical practice, research and training. London: Jessica Kingsley Publishers.

Wosch, T., & Wigram, T. (Eds.). (2007). Microanalysis in music therapy: Methods, techniques and applications for clinicians, researchers, educators and students. London: Jessica Kingsley Publishers.