Approaches: An Interdisciplinary Journal of Music Therapy
Volume 18 | 2026 | https://doi.org/10.56883/aijmt.2026.601

ARTICLE OPEN ACCESS

Music therapy works… but how do music therapists assess it? Experiences from practitioners

Shun Ting Seren Tang[1]

[1] Independent scholar, UK

Abstract

The interest in music therapy assessment arises out of the profession's need to evidence effectiveness, a priority that has become even more relevant in today's economic climate. Considerable debate surrounds how best to conduct assessments, reflecting differing epistemological orientations within music therapy. A review of the literature on assessments in the fields of music therapy, research methodology, psychology, and international development reveals varying perspectives on what constitutes credible knowledge. Given the author's background in monitoring and evaluation in international development, this research project seeks to document practitioners' experience with using or developing assessment measures. Three music therapists with an experience of developing or adapting assessment measures were interviewed, and the data were analysed using thematic analysis. It was found that a wide range of criteria beyond notions of reliability and validity were considered to ensure the tool's credibility. Faced with pragmatic concerns, therapists were acutely aware that information can be lost during assessment, yet they navigated this creatively to develop a "good enough" tool that aligns with the important values in music therapy. Therapists also reported benefiting from the process, gaining clarity around their work and reflexivity around evidence. This study highlights the need for more documentation of practitioners' experiences with assessment in the field.

Keywords

music therapy assessment, assessment measures, practitioner experiences, credible knowledge, monitoring and evaluation, reliability and validity, practice-based evidence, assessment tool development, reflexivity

Received 4 April 2025; Accepted 24 August 2025; Published 27 January 2026

Editor: Giorgos Tsiris; **Reviewers:** Anonymous; **Language consultants:** Jonathan (Jaytee) Tang (English), Effrosyni Efthymiou (Greek); **Copyeditor:** Anna G. Castells

Background

> "Music therapists frequently find themselves in a creative tension between the art of music and the science of therapy." – Anne Lipe (2015, p. 76)

The quote above illustrates the long-standing tensions surrounding music therapy assessment. This probably began with the development of modern music therapy as a practice, discipline, and

Publisher: Queen Margaret University Edinburgh | Centre for Person-centred Practice Research

profession. As Ansdell (2014) explains, pre-modern notions of music as a healing force eventually gave way to the more narrowly framed view of music therapy as a component of medical or psychological treatment. As the young profession sought to establish itself in the medical community, there was an awareness of the need to demonstrate its effectiveness (Solomon, 1985; Wigram, 2006). In the UK, the current economic climate pushes the drive to demonstrate impact even more, particularly because music therapy is a relatively small and less frequently commissioned profession within the Allied Health Professions (Kirkwood, 2023). Where the drive for evidence is coming from a managerial perspective, "documentable evidence" is often favoured (DeNora, 2006, p. 83). DeNora (2006) also notes there are "evidentiary hierarchies" (p. 86), where experimental designs such as randomised controlled trials are seen as more prestigious than qualitative studies. As a result, practitioners face an ongoing challenge in meeting demands for recognised forms of evidence while still maintaining the integrity of music therapy practice (Ansdell, 2006).

With a background in monitoring and evaluation in international development, I am keen on approaches to communicating the usefulness of certain interventions. Assessing improvisational music making in the context of therapy is similar to thinking about monitoring and evaluation systems – personal and social change are both complex processes, involving multiple stakeholders with varied needs.

Working in the frontline of development projects also makes me excited about practical knowledge generated on the ground. I am interested in how research, theory, and practice "enrich, inspire and emerge from one another, and act in tandem to one another" (Ansdell & Pavlicevic, 2001, p. 21). A range of music therapy assessment tools across settings and client groups have been discussed in the literature (e.g., Cripps et al., 2016; Lipe, 2015). The interest in how practitioners make use of these resources – or even, devise their own measures – prompts me to select this topic for my trainee research project.

Looking at the wider picture, the consideration of music therapy assessment has consequences on what counts as knowledge (evidence) and "good" knowledge (evidence; DeNora, 2006, p. 81). The creative tension embodied in this topic may lead us into examining our assumptions on "health, illness, effectiveness and ineffectiveness" (DeNora, 2006, p. 91), shedding light on alternative ways of framing and seeing.

Literature review

"Assessment measures" can be understood as those "relevant for assessing or comparing a person's functioning, symptoms or presenting features when they participate in a treatment or therapy" (Cripps et al., 2016, p. 5). They can take place in various points of time – the beginning, middle, and end of therapy, and can measure outcomes, process, or the client-therapist relationship (Cripps et al., 2016; Turry, 2018).

As mentioned, the interest in assessment in music therapy arises out of the profession's need to evidence effectiveness. Assessment measures can help us learn more about the clients, what their needs are, and to what extent these have been addressed (Spiro et al., 2020). As such, music therapists, clients, and family members may all have an interest in them. Assessment measures can also support therapists' self-reflection and development (Kirkwood, 2023).

Putting a square peg in a round hole: The mismatch between ways of assessment and how music therapy works

While there is little controversy around *why* assessment measures are needed for music therapy, *how* they should be carried out is a subject of much debate. Scholars such as DeNora (2006) and Ansdell (2006) argue that where forms of evidence considered to be more prestigious – such as experimental designs with controls as in randomised controlled trials or Evidence Based Practice – are used to measure and represent music therapy, lots of information is missed. Both point to music therapy as a mode of activity that is a form of human cultural interaction, with benefits beyond physiological terms. In such interactions, both the clients' and music therapists' musical contributions are part of the effects of music therapy.

Looking into more recent literature, it seems that DeNora's and Ansdell's observations are still very relevant. As Lipe (2015) discusses music-based assessment, she seems to limit the discussion to forms of measurement of a person's musicality as a construct (a hypothetical psychological attribute that cannot be measured directly, so its presence must be inferred through observed behaviours), with experimental designs involving musical protocols (e.g., tasks, stimuli, and cues to elicit potential strengths).

The differing views on what are appropriate assessment measures for music therapy may reflect a difference in an individual's epistemological orientation to music therapy. Drawing on his work in the ethnomusicology of autism, Bakan (2014) notes that music therapy is often described as treatment-focused, with the expectation that the client should change and therapy aims primarily at reducing symptoms and promoting behavioural and functional norms. He argues that this perspective should be balanced with perspectives of autistic self-advocacy, neurodiversity, and disability studies, highlighting the importance of responding supportively to the creative initiatives of autistic individuals. A similar view is found in Verney and Ansdell's (2010) discussion of how the Nordoff-Robbins music therapy approach views pathology – instead of circumventing it, the approach works in and with it. The therapist works towards helping clients to stand in a different relationship to their pathology, experiencing something different from how they habitually experience themselves.

Assessment tools which bear an assumption of the more knowledgeable therapist treating an individual who needs to be cured would not be able to capture the above understanding of music therapy. This is reflected in a recent debate on measuring the efficacy of music therapy with autistic children. Turry (2018, p. 87) responds to a randomised clinical trial, commenting on the "exclusive focus on symptom reduction rather than examining treatment effects for functional gains and quality of life", and the lack of process measures, such as client engagement with music-making and the quality of the therapist–client relationship. The limitation of clinical trials is also recognised by the authors of the said trial: "Essentially, a trial asks about the effects of offering an intervention (where not all will engage equally), as opposed to the effects of engaging with it" (Gold & Bieleninik, 2018, p. 91).

Efforts have been carried out to capture the relational nature of music therapy, that clients are resourceful and that the process is shaped by the clients rather than just the therapist's preexisting goals. Examples include the Child-Therapist Relationship in Coactive Musical Experience scale and the Musical Communicativeness scale developed by Nordoff and Robbins (1977). These scales were

developed to adequately assess the range of clinical music experience, while not hampering both objective perception and creative freedom (Nordoff & Robbins, 1977). Turry (2018) suggests the rating scales focus on the questions of "How are the therapist and client engaged in music-making?" and "How does increased musical engagement lead to overall improvement in the child's life?" (p. 88). Ansdell and Pavlicevic (2010) summarised the way in which Nordoff and Robbins developed their approach to music therapy as from a "gentle empiricism" tradition, a qualitative stance which is pragmatic and empirical. Features such as allowing the emergent phenomena to show themselves, building a collection of exemplary cases for demonstration and comparison, and searching for the varying circumstances in which the same phenomenon occurs have given rigour to the research. Further, a process measure to study the quality of musical engagement is currently in development (Nordoff-Robbins Center for Music Therapy, 2023).

Differing views on credibility of evidence

Drawing on my background in international development and to put the discussion within a wider context, this section examines how credibility of evidence is considered in the fields of research methodology, psychology, and international development.

1. Approaching the credibility of experimental designs critically

The evidentiary hierarchies noted in the previous section may be due to the perception that qualitative research may not possess adequate credibility as other forms of evidence. Considerations such as reliability (the ability of a measure to give consistent results under similar circumstances), validity (the tool measures what it intends to measure), and standardised tools are mentioned (Lipe, 2015; Spiro et al., 2020; Wigram, 2006). Resources listing a range of assessment measures, such as Lipe (2015) and Cripps et al. (2016) exclude "informal interviews" and include only "formal assessments" (where statistical methods are used to examine a tool's reliability and validity).

In psychology, the concept of "ecological validity" may offer a critical reflection on experimental designs. It refers to "the relation between real-world phenomena and the investigation of these phenomena in experimental contexts", and examines the representativeness and naturalness of the nature of the setting, the stimuli, and the response (Schmuckler, 2001, p. 420). The tension between standardising the intervention (stimuli) and having the intervention as close to as it would be in real life is observed in efforts to develop assessment tools for music therapy (Bell et al., 2014) and musical engagement (Ockelford & Welch, 2012). Both studies try to standardise the intervention in forms such as format, venue, and duration. Bell et al. (2014) admit that "each [music therapy] session was unique" (p. 63). Ockelford and Welch (2012) further identify factors making control of the intervention difficult. These include participants becoming increasingly familiar with the materials over time and the researcher developing a deeper understanding of the participants, enabling more effective interactions.

2. Broadening the concept of validity

Lincoln and Guba (2005) suggest validity is essentially a "goodness criteria" (p. 199). They argue that the central question researchers should be asking themselves should be: "Are these findings

sufficiently authentic… that I may trust myself in acting on their implications?" (p. 205) Relating to this concern are concepts such as fairness (balance of all stakeholders' views in relation to the researcher's own biases), positionality, critical subjectivity, and reciprocity (rather than a hierarchical research relationship). They further suggest that validity can be achieved not just by an application of certain methods, but through being interpretively rigorous, such as "defensible reasoning, plausible alongside some other reality that is known to author and reader" (p. 205).

A similar view is found in international development. Chambers (2015) points out that rigour, understood as a lack of bias, is associated with a paradigm of controllable and predictable physical things; where in development practice, one works in conditions of complexity, emergence, nonlinearity and unpredictability. Accordingly, he suggests adopting the "canons of inclusive rigour for complexity": eclectic methodological pluralism, improvisation and innovation, adaptive iteration, triangulation, plural perspectives, and being open, alert and inquisitive (Chambers, 2015, pp. 328-329). Chambers' canons are similar to Lincon and Guba's (2005) goodness criteria, with the exception of the inclusion of practical concerns. He adds the criterion of "optimal ignorance and appropriate imprecision", entailing not finding out more than is needed (Chambers, 2015, p. 329). Where "plural perspectives" are concerned, some music therapy literature similarly argues for the inclusion of multiple views such as service users' experiential knowledge of music therapy in statutory mental health service evaluations (McCaffrey, 2018), and voices of parents, siblings, teachers for certain autistic people who have yet found any alternative mode of voice to communicate their views (Bakan, 2014).

3. Addressing concerns around positive bias

Qualitative assessments may be perceived as idiosyncratic and that "there's a tendency to pick on the positives" (Procter & Tyrer, 2006, p. 79). Chambers (2015) addresses this specifically, arguing for the importance of "overt, transparent and self-critical reflexivity to recognise and offset biases" (p. 333). Such reflexivity has both a personal aspect (reflecting on one's own framings, categories, and mindset) and a methodological aspect (demanding critical examination and exposure of potential distortions) . Similarly, Gattino (2023) argues that the subjective nature of interpreting and assigning scores cannot be completely eliminated. However, therapists can help address this challenge by clearly defining the assessment tool's constructs, domains, dimensions, and the contents represented by its items.

Limited documentation of practitioners' experience

The incorporation of multiple perspectives can also include those of practitioners. In the existing literature on assessment in music therapy, the practitioners' experience of using, adapting, or developing assessment measures seems to be less documented. For instance, Lipe (2015) and Cripps et al. (2016) only consider professional documents and existing literature. Where music therapists were involved in a multi-disciplinary team in developing assessment tools, the literature tends to focus on the measures rather than the therapists' reflections of the process (e.g., Bell et al., 2014; Ockelford & Welch, 2012).

Research questions

Accordingly, the research considers the following questions relating to practitioners' experience of developing music therapy assessment tools:

1. What are the principles that guide the development of such practices?
2. What are the practitioners' reflections from the experience?

Methodology

Methodological tradition

While working in international development, I developed an interest in research methodologies which challenge knowledge generated by "instrumental rationality" (Fals Borda, 2001). Even when such knowledge brings technological advances, it has limits in findings solutions to complex social issues if it excludes the perspectives of people with lived experience. I am particularly interested in research methods which take into account subjective accounts of participants' lived experiences.

Accordingly, a constructivist paradigm and a qualitative methodology were chosen, focusing on the "entire phenomena" (e.g., events, experiences, materials, and persons) instead of stimulus or response variables (Bruscia, 1995). The qualitative methodology would suit well in documenting practitioners' experiences of moulding appropriately shaped holes for the music therapy "square peg," as its open-endedness would allow me to capture not only the known unknowns but also potentially the unknown unknowns, and bringing in the practitioners' subjective perceptions and values which may help deconstruct and reconstruct concepts. While the timeframe and scale of this master's project led me to focus on qualitative interviews, I was inspired by methodologies such as Participatory Action Research, which centre knowledge generated by lived experience.

Data collection

Three semi-structured interviews were conducted with music therapists from Nordoff and Robbins. As the present project involves interviewing music therapists, the fundamental ethical principles of beneficence or no harm, honesty, freedom and autonomy, and privacy (Maranto, 1995) were discussed in an ethical clearance process involving the Nordoff Robbins Research Ethics Committee. Ethical approval was granted in March 2024.

Given practical constraints and access to Nordoff and Robbins music therapists through my educational institution, these therapists were approached for this project. An open email invitation was sent to all Nordoff and Robbins music therapists in March 2024. No responses were received, possibly reflecting a perception among practitioners that their experience does not constitute "assessment." Consequently, I also asked my tutors for recommendations. Individual invitations were extended to eight music therapists; three of them agreed to be interviewed.

The interviews took place online in April 2024. Prior to the interviews, participants were encouraged to reflect on their past practices of assessing music therapy under four specific questions, namely:

1. What are the circumstances under which such practices are developed?
2. What are the principles that guide the development of such practices?
3. How are such practices conducted?
4. What are the practitioners' reflections from the experience?

The researcher followed a more detailed interview guide (see Appendix) in the interviews. For the two interviewees who had developed their own tools, electronic versions of these tools were shared with the researcher.

Participants

All interviewees were trained in the Nordoff Robbins approach to music therapy in the UK. This approach is often described as music-centred, as it emphasises the convergence of musical and personal processes, the intrinsic value of musical experience, and the use of music as an autonomous clinical force (Aigen, 2014). All interviewees have practised for more than 10 years in the UK and have worked in a range of settings. Interviewee 1 developed two tools – one for neurorehabilitation and another for use in a special school – both of which were utilised by the therapist. Interviewee 2 co-developed a tool for a special school with another music therapist working in the same setting and both used it in their practice. None of these tools were published. Interviewee 3 did not develop an assessment tool, but had experience of being involved in the process while working in an early intervention centre for children aged 0-5.

Data analysis

Audio recordings and transcripts were produced for all interviews, and the transcripts were edited for brevity and clarity. Thematic analysis – a method for "identifying, analysing and reporting patterns (themes) with data" (Braun & Clarke, 2006, p. 79) – was conducted. As the data was coded, I tried to give full and equal attention to each data item, code for as many potential themes as possible, and not try to smooth out any inconsistencies (Braun & Clarke, 2006, p. 89). I tried to follow the principles of gentle empiricism in the coding process, allowing the emergent phenomena to show themselves, instead of trying to fit the data into a preexisting coding frame (Ansdell & Pavlicevic, 2010). I noticed the initial codes were influenced by the literature and a seminar on assessment in the training course, leading to quite a body of data being coded under "usability", "validity", and "reliability". As I reread the transcripts, I began to identify new codes such as "generalisability," "representativeness," and "sensitivity."

The codes were then analysed to generate themes and sub-themes. Attention was paid to the "story" that each theme tells in relation to the research questions, such that there is not too much overlap between themes (Braun & Clarke, 2006). Themes were chosen not by the number of occurrences across the interviewees, but rather on whether it captures something important in relation to the overall research question.

Findings

The table below gives an overview of the four major findings from this research:

Theme 1: Goodness criteria and practical concerns were jointly considered as interviewees aimed for a good enough tool.	This theme discusses practices employed for ensuring the credibility of the assessment tools, such as statistical validation, critical reflexivity, representativeness, and sensitivity towards small changes. References to practical concerns such as usability, time, and cost further suggests the tool development process is more of finding a good enough tool rather than the best tool.
Theme 2: What is "good enough" is influenced by local needs.	This theme presents how perceptions of what is good enough for managers and people at the local level are further factors shaping the design of assessment tools.
Theme 3: Challenges and opportunities arising from developing assessment tools.	Using examples of measuring emotions and capturing the ripple effect of music therapy, this theme considers the mismatch between assessment methods and the music therapy process. Nevertheless, the tool development process could also become an opportunity for therapists to highlight alternative values or ways of seeing to a treatment-based model of music therapy.
Theme 4: How the process of developing and using assessment tools helped the therapists.	This final theme investigates how assessment tools helped music therapists clarify and communicate the value of their work, and gain reflexivity around evidence.

Table 1: Overview of findings

Theme 1: Goodness criteria and practical concerns were jointly considered as interviewees aimed for a good enough tool

The first finding highlights a similar structure across all four assessment tools. Interviewee 1 developed one tool for neurorehabilitation and another for use in a special school. Interviewee 2 created a tool for a special school, while Interviewee 3 contributed to the development of a tool for an early intervention centre. To protect the confidentiality of the interviewees and respect the intellectual property of their work or that of their organisations, specific details of the tools are not discussed in this paper. However, it is noted that all tools share a structure based on "dimensions" and "levels" – terms chosen by the researcher for consistency. "Dimensions" concern the areas to be rated, which ranges from 6 to 26 across the tools. On "levels", all tools use ordinal scales, and numbers are used in conjunction with word descriptors. Levels range from 4 to 5 across the tools.

The literature review discusses that a "good" assessment tool is viewed by some as *being "scientific" and statistically validated*. This view is echoed in the neurorehabilitation case, where the consultant of the unit suggested for the tool to be "*validated*" (Interviewee 1). The findings further suggest that inter-rater reliability, which refers to the consistency among multiple raters using the same tool, is considered a criterion of goodness:

> We did a focus group with Nordoff-Robbins music therapists to give a fuller description of the dimensions and the levels, with some numbers… The process was about making the same scale more usable by other people […] The descriptions of the dimensions were initially open-ended. The consultant and the musical psychology researcher were keen for me to narrow them down, specifically for the purpose of inter-rater reliability, to define things more. So, I ended up with things like 'taking more than 50% of opportunities to initiate'. (Interviewee 1)

However, it appears that there are goodness criteria that are beyond statistical validation. All interviewees mentioned criteria akin to "adaptive iteration" and "plural perspectives," as discussed in the international development literature (Chambers, 2015, pp. 328-329). Specifically, interviewees talked about how the tool development processes were participatory and iterative, incorporating multiple perspectives and preexisting frameworks in a particular setting. For instance, in a special school for children, interviewee 2 worked with another music therapist in the same setting and referenced the performance attainment target scale (for pupils aged 5-16 with special educational needs), school documents, other external tools used by teachers, the Nordoff-Robbins Rating Scales, and three other scales from another therapy discipline. Then, they checked with each other "to see if either of us have forgotten something, or place slightly higher reliance on one particular dimension."

Further, interviewee 1 tried to be led by data as much as possible and be slow in making inferences on what is "good" and "progress". This is similar to Lincoln and Guba's (2005) notion of *critical subjectivity*, and Chambers' (2015, p. 333) "overt, transparent and self-critical reflexivity":

> I was trying to codify my own subjectivity…, the [dimensions] were chosen to formalise my judgment on how music therapy has gone with each [individual I work with]. It depends on me being honest and as objective as possible […] The [music psychology] researcher's question of 'How do you know' was really helpful… [The researcher] was constantly pushing me towards what I could actually see and identify that made me judge what I judged. (Interviewee 1)

Interviewees suggested four further criteria of making a good assessment tool. The first concerns *generalisability*, the extent to which the tool can meaningfully be applied to all clients:

> We tried to find categories that would apply across a broad range of abilities, to reflect the motivation of a child with profound and multiple learning disability (PMLD) to engage musically. For example, it might mean they are looking intently, and they are breathing really hard to try and provide a sound. (Interviewee 2)

Second, the tool should be *representative* of the work of the entire music therapy service and the whole organisation. Interviewee 3 mentioned that to design a tool for the early intervention centre, the external consultant held an initial meeting with different members of staff, such as the fundraiser, the therapists, and home outreach workers. The goal was to have a tool that would cover all aspects of the charity's work.

Third, interviewee 2 mentioned that the dimensions were chosen partly based on what was observed empirically, in line with the *"gentle empiricism"* tradition (Ansdell & Pavlicevic, 2010). Based on the observation that, over time, children "were making more obvious choices, perhaps about the instruments that they were using, or what they wanted to do, whether they wanted to leave or carry on", the dimension of "Communicates clear choices" was included.

Fourth, interviewees working in special schools stressed that the tool should be *sensitive* enough to capture the smallest changes. Ockelford and Welch (2012) similarly note the difficulty of developing a tool refined enough to reflect changes for children with PMLD, "since, even in a 12-month period, it seemed likely that they would make only tiny increments of progress" (p. 20):

> At some point, the [dimension] of 'communication' was changed to 'expressive range', and 'learning skills' was changed to 'cognitive'. Because a lot of students in this school does not communicate verbally, or even intentionally. A lot of them do not engage in formal learning. So, for some students, an increase in expressive range is actually an important part of the way they communicate. (Interviewee 1; special needs school setting)

Considerations of goodness criteria were found to be balanced against *practical concerns*. All interviewees mentioned the importance of usability of the tool. First, there is an aspect of how easy it is for the rater to use. For instance, interviewee 2 noted: "we extracted key aspects from the original [Nordoff-Robbins Rating] Scales, such that people familiar with the original scales could recognise that, but it is simpler, less wordy."

Usability was also placed within an understanding that in reality, it is often the therapists themselves who carry out assessments; the therapist-rater should be spending most of the time delivering music therapy. As interviewee 2 noted: "we did not want to spend hours going through the videos to be able to rate… It was [also] in the interest of the headteacher that we did not spend our time writing reports."

Usability also has a user-facing dimension. Interviewees talked about the efforts made to give a clear presentation of outcomes:

> The spreadsheet contains a function, so that when we enter a number, the cell will automatically turn that into red, amber, yellow or green. The headteacher looked at it and thought it was good to potentially see developments quickly from the colours. (Interviewee 2)

Finally, it is worth noting that even in a medical setting, pragmatic concerns about time and cost can take precedence over the insistence on statistically convincing findings:

> I selected a few short extracts, and a group of music therapists rated each one. We got all the data, and that is as far as it got. The next stage would have been doing statistical tests on the data. A number of people said they would do it, and they never did… It felt like there was a lot of interest from a lot of people in producing the scale, up to the point where something like statistical analysis needed doing, and then nobody really wanted to do it. (Interviewee 1; neurorehabilitation unit setting)

Taken together, theme 1 presents a range of considerations in assessment tool development. The findings suggest a wide practice for ensuring the credibility of the tools, such as incorporating multiple perspectives, critical reflexivity, representativeness, and sensitivity towards small changes. All interviewees further stressed the importance of pragmatic concerns. This points to that in reality, the process can be more of finding a good enough tool rather than the best tool.

Theme 2: What is "good enough" is influenced by local needs

Apart from thinking along the lines of goodness criteria, this research finds that perceptions on what is good enough for the managers and people at the local level shaped the design of the assessment tools. All interviewees mentioned getting or justifying existing funding as the drive for developing assessment tools. While the music therapists seemed to find it relatively easy to perceive the needs of the people whom they regularly interacted with, it was less clear as to what was needed at the managerial level. At times, there were also uncertainties around who were the people asking for evidence:

> The headteacher was put under pressure to justify the use of pupil premium funding. It could be a government requirement for the local authority to justify expenditure of pupil premium just generally within schools. It may also be internally, from the Board of Governors. (Interviewee 2)

Further, interviewees reported that written reports produced by themselves addressing changes in dimensions of interest to the settings were good enough to satisfy local needs. Even in a medical setting, staff members showed more interest in written reports than numbers generated from a specifically designed assessment tool:

> I produced all these numbers from the scale. I don't know if anybody ever looked at it or taken any notice of it. At the same time, I would write a report for every patient's discharge, and notes for the ward rounds… Thinking of the 'dimensions' in the scale was very helpful, in thinking about what kinds of things to talk about qualitatively… Those did get read, and people would often get back to me. (Interviewee 1; neurorehabilitation unit setting)

Word of mouth and the direct experience of participating in music therapy were also found to be important to convince people at the local level:

> The headteacher of the school said he valued music therapy very highly. I said why, you've never seen it. He said, my staff tell me… One of the things that people really notice was that, when I went to the classroom to pick up the students… as soon as they see me, they came running out and grab my hand and take me to the session. That's almost the main piece of evidence that people would notice. (Interviewee 1; special needs school setting)

The above suggests a similar view to when Ansdell (2006) recalled his colleague Rachel Verney's experience in setting up music therapy work: "Her experience is that local demands mostly trump central criteria; that showing what music therapy could do for a particular place and client group is as important as 'proving' it in some abstract sense" (Ansdell, 2006, p. 97).

That said, this research finds a case where central and local needs aligned. Interviewee 2 mentioned a new system that was developed in the school to streamline the process of reporting developments of the children and the management communicated to staff what were needed. The interviewee was able to spend less time writing reports and showed to the school what happened in the sessions directly through uploading video extracts: "Our notes became the school's notes, which anybody within the school has access to; everybody can see what we have been doing through these snippets of videos."

Summing up, there seemed to be a general lack of clarity as to what was exactly needed from the management. In some cases, this may lead to the therapists doing more than is needed. In contrast, therapists had a much better understanding of the needs of the people whom they regularly interacted with, which helped them tailor their communications effectively.

Theme 3: Challenges and opportunities arising from developing assessment tools

3.1 The mismatch between assessment methods and the music therapy process

As mentioned in the literature review, DeNora (2006) and Ansdell (2006) discuss the potential loss of information important to music therapy when assessment tools are used. Others highlight that the treatment-directed paradigm of the profession does not truly represent the views of all music therapists (e.g., Bakan, 2014). Findings suggest that interviewees were aware of the limitations of assessment tools in aligning with their views of the therapy process, particularly that the tools carry an assumption that the people they work with needs to be treated:

> It always was hard to judge somebody on how they're behaving. How they interact with you one week may be very different from another week. If you try to plot them on a scale, are you taking into account the whole person?… It felt as if there was quite a tricky juxtaposition between outcome measures and allowing full self-expression and acceptance of a person as they are. (Interviewee 3)

Two specific examples were given. First, all interviewees talked about the difficulty of measuring emotions. It comes with the uncertainty around defining progress in terms of changes in emotion states and the difficulty of naming the emotions for those who may not be able to self-report. As interviewee 1 said: "I decided that 'emotions' could not be represented in an ordinal scale, because you could not say moving from one to another was necessarily progress or the opposite."

The above views are likely because all interviewees are Nordoff-Robbins trained music therapists. The Nordoff-Robbins approach emphasises accepting and meeting a person's emotional state, then matching, accompanying, and enhancing the person's expression, so it tends towards a primary experience of intercommunication (Nordoff & Robbins, 1977). Procter (2016) further argues

that music therapy is not about emotional expression in and of itself; as emotions are part of people's experience of being themselves, the enhancement of expressivity when one engages in music-making can intersubjectively reshape one's experience of one's way of being, thus discovering new ways of being as a person, and as a person with another person. This view is echoed in the tool developed by interviewee 1 in a neurorehabilitation unit setting, as "expressive range", understood as "range of expressive components employed in music-making (e.g., tempo, pitch range, volume, tone)", was included. It is worth noting that other approaches to music therapy may take a different view. For instance, Thaut and Wheeler (2010) discuss the role of music in influencing and modifying affective states, bringing about behavioural learning and change.

A second example of what may be missed in assessment is the ripple effect of music therapy, understood as impact that "goes beyond the individual client, to reach families, carers, as well as other staff members" (Tsiris et al., 2018, p. 6). In the case for children with brain injuries or other disabilities, music therapy may help families process the loss and find new ways to be with their children, but this ripple effect may not be captured by assessment tools which focus on the child:

> These are families which have just been given a diagnosis which is life-changing to them… I think the music therapy is about finding new ways to be with your child and interact with them… If you're only working with someone for half an hour a week, there's only so much realistically you can actually do. But if you're empowering the parents to believe in the child's ability, then you're helping to set up for things to be happening outside… In the end the tool does not have 'family life' [as a dimension]; it would come under communication and social interactions. (Interviewee 3)

3.2 The tool development process provided an opportunity for therapists to advocate new ways of seeing

It is found that the tool development process can become an opportunity for the therapists to highlight what music therapy can uniquely offer in a setting vis-à-vis other activities and what values it advocates. In a medical setting, the therapist suggested that alternative ways of seeing people with brain injuries are possible. For example, interviewee 1 admitted the structure of the assessment tool was similar to the Functional Independence Measure (FIM) and Functional Assessment Measure (FAM), which was widely used in the unit and in neurorehabilitation more generally, and suggested there was a compromise:

> I was a bit uncomfortable to use an ordinal scale of levels of limitation. It was similar to the FIM+FAM scale. I do not think people without brain injury are unlimited in music-making… I do not think it is true for any of us.

This view made its way to the text of the tool, which reads:

> Appropriate use […] It covers a range of severities from minimally responsive patients to patients with only a mild disability, and could also apply to a non-brain injured population. It would not necessarily be expected that a non-brain injured person would score 'no limitation' on all the dimensions.

In educational settings, therapists were found showcasing what the child *can* do, since "negative behaviours" can be viewed positively when placed in the context of interactive music-making, or that what the child is capable of may not be so apparent in other environments. This can be important for both staff members and families:

> The Head would say… the boundaries and barriers did not exist in music therapy in the way that it did in other settings… The Head also said that the possibility for children with PMLD to demonstrate learning and development was often clearer in music therapy at a very base level. If the children want to get a college place, they would need to demonstrate they are still learning. (Interviewee 2)

Summing up, it is found that the interviewees were well aware that the very act of developing an assessment tool with dimensions and levels would inevitably involve degrees of simplifying the music therapy process. The difficulty of incorporating how music therapy views emotions and its ripple effects were given as examples. Notwithstanding this, the tool development process could become an opportunity for therapists to highlight alternative values or ways of seeing compared to a treatment-based model of music therapy, implicitly redressing the therapist-client power imbalance. Therapists were found highlighting the capabilities of the people they work with and recognising that they are active agents for change.

Theme 4: How the process of developing and using assessment tools helped the therapists

4.1The process aided therapists to clarify the value of their work

Even in situations where the tools were not eventually used, the process helped the therapist "in thinking what is important and relevant, and what is the common ground [with other disciplines]" (Interviewee 1). In the case where the assessment tool was used regularly in assessments and reviews, the interviewee talked about how the tool helped in refocusing longer-term work, clarifying the therapeutic direction:

> If you've been working with someone for a year, it would be really helpful to just have a refocus – what am I doing, why am I doing this, what are our goals… Goals such as 'to explore self-expression' can be quite airy fairy – what does that actually mean? If I'm using a more formalised, specific tool, it can help you to focus on what you mean by exploring self-expression. (Interviewee 3)

4.2 The process helped therapists to communicate the value of their work

All interviewees mentioned how thinking around and/or using assessment tools enhanced their communication with the wider team when working with the same clients. That can be in the form of the therapists presenting the importance and relevance of music therapy in a clearer framework to the

team, being more consistent in language when more than one music therapist was working in the same setting, or getting inspiration from the team to refocus one's work.

A further finding is that, as one case shows, the assessment tool can become part of an existing participatory assessment and monitoring process, incorporating perspectives other than the therapists. Talking about the work at an early intervention centre, interviewee 3 mentioned that the tool was regularly used in assessment and reviews of the children with the parents. Should they wish to be involved, families can contribute their views on what they believe a child's starting point is. The assessment tool also provides focal points when parents and staff discuss future goals for the children. Interviewee 3 saw great value in this process:

> When [the children are] that young, so much can change and so much is unknown when they first arrived. They may have doctors telling them they can't do this and others. When they are starting, you can't actually tell what a child will do, so, your goal might have been down here initially; but actually on review you might push it a bit more because it ends up they are capable of more. (Interviewee 3)

4.3 The therapist gained reflexivity around evidence

Finally, interviewee 1 showed a "reflexive and questioning stance towards the entire range of evidence and impact-related endeavours in music therapy" (Tsiris et al., 2018, p. 24), as they talked about how the process helped them think about evidence more generally:

> We music therapists often think we are very different from all these other disciplines which are very evidenced based. Actually, realising that disciplines like medicine, physiotherapy, speech and language therapy… they are actually struggling in exactly the same way as we do; it is really hard for them to make evidence. There are lots of questions and doubts over the way they produce it. That was really useful. It was not like that we have no evidence, and everybody else has loads of evidence. (Interviewee 1)

A similar view is expressed by Sapiro et al. (2020), as they argue that "[m]usic therapy is not the only field in which questions around sample size, statistical methods and reporting have arisen" (p. 19).

Conclusion

Drawing on practitioners' experience in music therapy assessment, this research identifies a variety of practices of ensuring the credibility of the tool that extend beyond statistical measures. Examples include participatory and iterative processes, the incorporation of multiple perspectives, and practitioners' critical reflexivity.

The research also highlights a broader understanding of credibility as a form of 'goodness criteria', those that enable practitioners to act with confidence on the implications of assessment

findings. Within this expanded view, concepts such as generalisability, representativeness, and sensitivity towards small changes are recognised as important considerations when evaluating assessment tools.

Overall, this research affirms the value of learning from the practitioners. From their stories of what happened on the ground, aspects such as practical concerns and perceptions enrich the picture of music therapy assessment. The process of assessment tool development were dynamic stories of idealism meeting pragmaticism. The music therapists were found innovating along the way, as they were met with practical challenges such as uncertainties around managerial needs and time constraints, while trying hard to make the tool useful, credible, and embodying the important values of music therapy.

Following from the findings that therapists can be uncertain around what is expected from the funders and that they aspire to redress the therapist-client power imbalance, future research can consider interviewing the clients and those who make funding decisions. All interviewees reported gaining invaluable insights on the value of their work through their assessment tools. For some, they turned this exercise of developing assessment tools into an opportunity to advocate alternative ways of understanding health and effectiveness of music therapy. These findings strengthen the case for further research and documentation of practitioners' experiences in developing and using assessment tools.

Author information

Shun Ting Seren Tang (she/her)holds a Master of Music Therapy from Goldsmiths, University of London (Nordoff Robbins) and a Master of Philosophy in Development Studies from the University of Oxford. She is currently a music therapist working in London (UK), supporting adults experiencing mental health issues, children with special educational needs, under-5s and their families, and adults undergoing neuro-rehabilitation.

@ **serenmusictherapist@gmail.com**

Acknowledgments

This article has been condensed from a Master's degree research project carried out during training with Nordoff and Robbins. I would like to thank the three interviewees for generously sharing their wealth of knowledge and experience. I would also like to thank my three tutors for recommending suitable interviewees, making this project possible. Last but not least, I would like to express my gratitude to my research supervisor, Jacob Harrison, for not only suggesting pertinent literature and thus making the project more exciting, but also being such an attentive and encouraging supporter.

Author contributions

Shun Ting Seren Tang: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing - original draft.

Artificial Intelligence (AI) usage

No AI tools were used in this research project.

Conflict of interest

The author has no conflicts of interest to report.

Funding

No funding was received for this study.

Data availability statement

The data are not available due to the nature of the study and its research ethics approval.

References

Aigen, K. (2014). Music-centred dimensions of Nordoff-Robbins music therapy. *Music Therapy Perspectives, 32*(1), 18-29. https://doi.org/10.1093/mtp/miu006

Ansdell, G. (2006). Response to Tia DeNora. *British Journal of Music Therapy, 20*(2), 96-99. https://doi.org/10.1177/13594575060200020

Ansdell, G. (2014). *How music helps in music therapy and everyday life*. Ashgate Publishing Ltd.

Ansdell, G. & Pavlicevic, M. (2001). *Beginning research in the arts therapies*. Jessica Kingsley Publishers.

Bakan, M.B. (2014). Ethnomusicological perspectives on autism, neurodiversity, and music therapy. *Voices: A World Forum for Music Therapy*, *14*(3). https://doi.org/10.15845/voices.v14i3.799

Bell, A.P., Perry, R., Peng, M., & Miller, A.J. (2014). The Music Therapy Communication and Social Interaction Scale (MTCSI): Developing a new Nordoff-Robbins scale and examining interrater reliability. *Music Therapy Perspectives, 32*(1), 61–70. https://doi.org/10.1093/mtp/miu002

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

Bruscia, K. (1995). Differences between quantitative and qualitive research paradigms: Implications for music therapy. In B. Wheeler (Ed.), *Music therapy research: Quantitative and qualitative perspectives* (pp. 65-76). Barcelona Publishers.

Chambers, R. (2015). Inclusive rigour for complexity. *Journal of Development Effectiveness, 7*(3), 327-335. https://doi.org/10.1080/19439342.2015.1068356

Chaytor, N. & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13*(4), 181-197. https://doi.org/10.1023/B:NERV.0000009483.91468.fb

Cripps, C., Tsiris, G., & Spiro, N. (2016). Outcome measures in music therapy: A free online resource by the Nordoff Robbins Research Team. Nordoff Robbins. https://eresearch.qmu.ac.uk/bitstream/handle/20.500.12289/4429/4429.pdf?sequence=1&isAllowed=y

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. https://doi.org/10.1177/135945750602000203

Fals Borda, O. (2001). Participatory (action) research in social theory: Origins and challenges. In P. Reason & H. Bradbury (Eds.), *Handbook of action research: The concise paperback edition* (pp. 27-39). SAGE.

Gattino, G.S. (2023). A commentary on 'The Communication-Relationship Outcome Matrix (CROM): A tool for measuring communication outcomes in everyday music therapy practice' written by Jenny Kirkwood. *Approaches: An Interdisciplinary Journal of Music Therapy, 15*(2), 328-332. https://doi.org/10.56883/aijmt.2023.10

Gold, C. & Bieleninik, Ł. (2018). Authors' response. *Nordic Journal of Music Therapy, 27*(1), 90-92. https://doi.org/10.1080/08098131.2018.1398988

Kirkwood, J. (2023). The Communication-Relationship Outcomes Matrix (CROM): A tool for measuring communication outcomes in everyday music therapy practice. *Approaches: An Interdisciplinary Journal of Music Therapy, 15*(2), 297-327. https://doi.org/10.56883/aijmt.2023.9

Lincoln, Y.S. & Guba, E.G. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N.K. Denzin & Y.S. Lincolnn (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 191-216). SAGE.

Lipe, A. (2015). Music therapy assessment. In B. Wheeler (Ed.), *Music therapy handbook* (pp. 76-90). Guilford Press.

Maranto, C.D. (1995). Ethical precautions. In B. Wheeler (Ed.), *Music therapy research: Quantitative and qualitative perspectives* (pp. 79-93). Barcelona Publishers.

McCaffrey, T. (2018). Evaluating music therapy in adult mental health services: Tuning into service user perspectives. *Nordic Journal of Music Therapy, 27*(1), 28-43. https://doi.org/10.1080/08098131.2017.1372510

Nordoff, P., & Robbins, C. (1977). *Creative music therapy: A guide to fostering clinical musicianship*. Barcelona Publishers.

Nordoff-Robbins Center for Music Therapy (2023). Research. https://steinhardt.nyu.edu/nordoff/research#:~:text=The%20Music%20Engagement%20Scale%20(MES,music%20making%20during%20NRMT%20sessions

Ockelford, A. & Welch, G.F. (2012). Mapping musical development in learners with the most complex needs: The Sounds of Intent Project. In McPherson, G.E. & Welch, G.F. (Eds.), *The Oxford handbook of music education* (Vol. 2, pp. 11-30). https://doi.org/10.1093/oxfordhb/9780199928019.013.0002_update_001

Procter, S. (2016). Playing my feeling or feeling my playing? A music-centred perspective on 'emotional expression' in music therapy. In L. Konieczna-Nowak (Ed.), *Music therapy and emotional expression: A kaleidoscope of perspectives* (pp. 55-68). The Karol Szymanowski Academy of Music.

Procter, S. & Tyrer, P. (2006). Interview: Professor Peter Tyrer. *British Journal of Music Therapy, 20*(2), 76-80. https://doi.org/10.1177/135945750602000202

Schmuckler, M.A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, 2(4), 419–436. https://doi.org/10.1207/S15327078IN0204_02

Solomon, A.L. (1985). *A historical study of the National Association for Music Therapy, 1960-1980.* [Doctoral dissertation, University of Kansas].

Spiro, N., Tsiris, G. & Cripps, C. (2020). 'Sounds good, but... what is it?' An introduction to outcome measurement from a music therapy perspective. *Approaches: An Interdisciplinary Journal of Music Therapy, 12*(1), 8-29. https://doi.org/10.56883/aijmt.2020.193

Thaut, M.H. and Wheeler, B.L. (2010). Music therapy. In P. N. Juslin (Ed.), *Handbook of music and emotion: Theory, research, applications* (pp. 818-848). Oxford University Press.

Tsiris, G., Spiro, N. & Pavlicevic, M. (2018). Repositioning music therapy service evaluation: a case of five Nordoff-Robbins music therapy service evaluations in neuro-rehabilitation. *Nordic Journal of Music Therapy, 27*(1), 3-27. https://doi.org/10.1080/08098131.2016.1273966

Turry, A. (2018). Response to effects of improvisational music therapy vs. enhanced standard care on symptom severity among children with autism spectrum disorder: the TIME-A randomized clinical trial. *Nordic Journal of Music Therapy, 27*(1), 87-89. https://doi.org/10.1080/08098131.2017.1394902
Verney, R. & Ansdell, G. (2010). *Conversations on Nordoff-Robbins Music Therapy*. Barcelona Publishers.
Wigram, T. (2006). Response to Tia DeNora. *British Journal of Music Therapy, 20*(2), 93-96. https://doi.org/10.1177/13594575060200020

Appendix

Interview guide (Set 1)

1. How long have you been a music therapist? What settings have you found yourself working in?
2. Think of a time when you adapted or developed a music therapy assessment practice/tool/measure that is memorable to you. What was the background (e.g. the setting, population, who was driving it, for what purpose)?
3. As you work on the assessment practice, what were the initial challenges?
4. What were the philosophical perspectives / rationale guiding the development of such assessment practice? What existing practices gave you inspirations?
5. What was the process of developing the measure?
6. How does the assessment practice work (e.g. design, application, analysis)?
7. How was the experience like using the tool?
8. What are your reflections (good practices, things that you would do differently, things that surprised you) around developing and using the assessment practice (e.g. what were measured, what were missed, stakeholders' perceptions of good enough evidence, your trust and confidence in the findings, in what ways do assessment affect the therapeutic relationship and process)?
9. (If there is more than one memorable story, repeat questions 2-8. Can also ask: What are the major similarities in the stories?)

Interview guide (Set 2)

This interview guide was for interviewee with knowledge of how *others* develop/use assessment tools.

1. How long have you been a music therapist? What settings have you found yourself working in?
2. Think of a time when you find others adapting or developing a music therapy assessment practice/tool/measure that is memorable to you. What was the background (e.g. the setting, population, who was driving it, for what purpose, measuring what)?
3. What was the process of developing the measure (e.g. initial challenges, existing practices that gave inspirations)? In what ways were you involved?
4. How does the assessment practice work (e.g. design, application, analysis)?
5. What do you observe about the experience of using the tool? What worked well and what did not? How far does the assessment affect the therapeutic relationship and process?
6. How does music therapy fit into this? (e.g. How far the assessment practice measures what you think should be measured? How far are the process and outcome useful to you? Would you like to be involved in developing / using the tools, and why (not)?)

Ελληνική περίληψη | Greek abstract

Η μουσικοθεραπεία λειτουργεί… αλλά πώς την αξιολογούν οι μουσικοθεραπευτές; Οι εμπειρίες των επαγγελματιών

Shun Ting Seren Tang

Μετάφραση: Ευφροσύνη Ευθυμίου

Περίληψη

Το ενδιαφέρον για την αξιολόγηση στη μουσικοθεραπεία προκύπτει από την ανάγκη του επαγγέλματος να τεκμηριώσει την αποτελεσματικότητά του, μια προτεραιότητα που έχει γίνει ακόμη πιο σημαντική στο σημερινό οικονομικό περιβάλλον. Υπάρχει έντονη συζήτηση σχετικά με τον βέλτιστο τρόπο διεξαγωγής αξιολογήσεων, η οποία αντικατοπτρίζει τους διαφορετικούς επιστημολογικούς προσανατολισμούς εντός του χώρου της μουσικοθεραπείας. Μια ανασκόπηση της βιβλιογραφίας σχετικά με τις αξιολογήσεις στους τομείς της μουσικοθεραπείας, της ερευνητικής μεθοδολογίας, της ψυχολογίας και της διεθνούς ανάπτυξης αναδεικνύει ποικίλες αντιλήψεις για το τι συνιστά αξιόπιστη γνώση. Λαμβάνοντας υπόψη το υπόβαθρο της συγγραφέως στην παρακολούθηση και αξιολόγηση προγραμμάτων διεθνούς ανάπτυξης, η παρούσα ερευνητική εργασία αποσκοπεί στην καταγραφή της εμπειρίας των επαγγελματιών σχετικά με τη χρήση ή την ανάπτυξη εργαλείων αξιολόγησης. Πραγματοποιήθηκαν συνεντεύξεις με τρεις μουσικοθεραπευτές που είχαν εμπειρία στην ανάπτυξη ή την προσαρμογή εργαλείων αξιολόγησης, και τα δεδομένα αναλύθηκαν μέσω θεματικής ανάλυσης. Διαπιστώθηκε ότι ένα ευρύ φάσμα κριτηρίων, πέρα από τις έννοιες της αξιοπιστίας και της εγκυρότητας, εξετάστηκε ώστε να διασφαλιστεί η αξιοπιστία του εκάστοτε εργαλείου. Αντιμέτωποι με πρακτικές προκλήσεις, οι θεραπευτές είχαν βαθιά επίγνωση ότι πληροφορίες μπορεί να χαθούν κατά τη διαδικασία της αξιολόγησης· ωστόσο, διαχειρίστηκαν αυτή την πραγματικότητα δημιουργικά, αναπτύσσοντας ένα «αρκετά καλό» εργαλείο που να συνάδει με σημαντικές αξίες στη μουσικοθεραπείας. Οι θεραπευτές ανέφεραν επίσης ότι ωφελήθηκαν από τη διαδικασία αποκτώντας μεγαλύτερη σαφήνεια σχετικά με το έργο τους και αναστοχαστικότητα σε σχέση με την έννοια της τεκμηρίωσης. Αυτή η μελέτη αναδεικνύει την ανάγκη για περαιτέρω καταγραφή των εμπειριών των επαγγελματιών σχετικά με την αξιολόγηση στη μουσικοθεραπεία.

Λέξεις κλειδιά

μουσικοθεραπευτική αξιολόγηση, εργαλεία αξιολόγησης, εμπειρίες επαγγελματιών, αξιόπιστη γνώση, παρακολούθηση και αξιολόγηση, αξιοπιστία και εγκυρότητα, αποδεικτικά στοιχεία βασισμένα στην πρακτική, ανάπτυξη εργαλείων αξιολόγησης, αναστοχαστικότητα