



City Research Online

City St George's, University of London

Citation: Spiro, N., Harkness, A. C., Sanfilippo, K. R. & Schober, M. F. (2026). Do People See the Same Things When They Watch Video Clips from a Music Therapy Improvisation Session?. Music & Science, 9, doi: 10.1177/20592043261456906

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/38283/>

Link to published version: <https://doi.org/10.1177/20592043261456906>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Do People See the Same Things When They Watch Video Clips from a Music Therapy Improvisation Session?

Music & Science
Volume 9: 1–22
© The Author(s) 2026
DOI: 10.1177/20592043261456906
journals.sagepub.com/home/mns



Neta Spiro^{1,2} , A. C. Harkness³ , Katie Rose Sanfilippo⁴ and Michael F. Schober³

Abstract

To what extent do different people find the same moments in a music therapy improvisation session salient, and how similarly or differently do they interpret the same moments? Four half-hour sessions at a music therapy center between different trained music therapists and musically experienced first-time attendees were videorecorded. Immediately after the sessions each participant watched the video of their session individually and commented on the one to five moments that struck them as most noteworthy. For each session, two other participants—one trained music therapist and one other musically experienced observer—also watched the videos and commented on one to five noteworthy moments. An online questionnaire for each session was subsequently created that included all four participants' comments (anonymized and edited for ratability) and the corresponding video segments. Each participant individually rated the extent to which they agreed with all the anonymized comments about their session. Analyses focused on the extent to which different participants selected the same moments as worthy of commentary, their levels of endorsement of comments originally authored by participants in the different roles, and overall patterns of rating agreement between participants in the different roles. Overall levels of agreement were low, but across the four sessions the trained music therapists' judgments tended to be more aligned with each other's, with some asymmetries: Session music therapists endorsed statements by the commenting music therapists more, but not the other way around. There was also evidence for greater agreement between session outsiders (commenting music therapists and commenting observers) than between session participants (session music therapists and attendees). In fact, by some measures session attendees' interpretations of what happened were particularly different from everyone else's. The findings are not consistent with a view that everyone is likely to see the same thing in a video clip of a moment in music therapy.

Keywords

Improvisation, interpretation, intersubjectivity, listeners, music therapy, outsiders, overlapping cognition, participation, shared understanding, training

Received: December 1, 2025; accepted: May 5, 2026

Action Editor: Bonnie McConnell, Australian National University, School of Music.

Peer Review: Alison Talmage, University of Auckland.

Jacqueline Birnbaum, New York University, Nordoff–Robbins Center for Music Therapy.

Consider this brief clip of a music therapy improvisation, in which a music therapist at the piano and a first-time session attendee at the drums make music together (See Figure 1): What is the best way to characterize what happened here, beyond noting that two people are playing together? Did something interesting happen? Was one of the parties leading and the other following, and did this change over time? How well was each party listening to the other? How free (vs. constrained) was either party's music-making?

¹Centre for Performance Science, Royal College of Music, London, UK

²Department of Surgery and Cancer, Imperial College London, London, UK

³Department of Psychology, The New School for Social Research, New York, USA

⁴City St George's, University of London, London, UK

Corresponding Author:

Neta Spiro, Centre for Performance Science, Royal College of Music, London, UK.

Email: neta.spiro@rcm.ac.uk



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

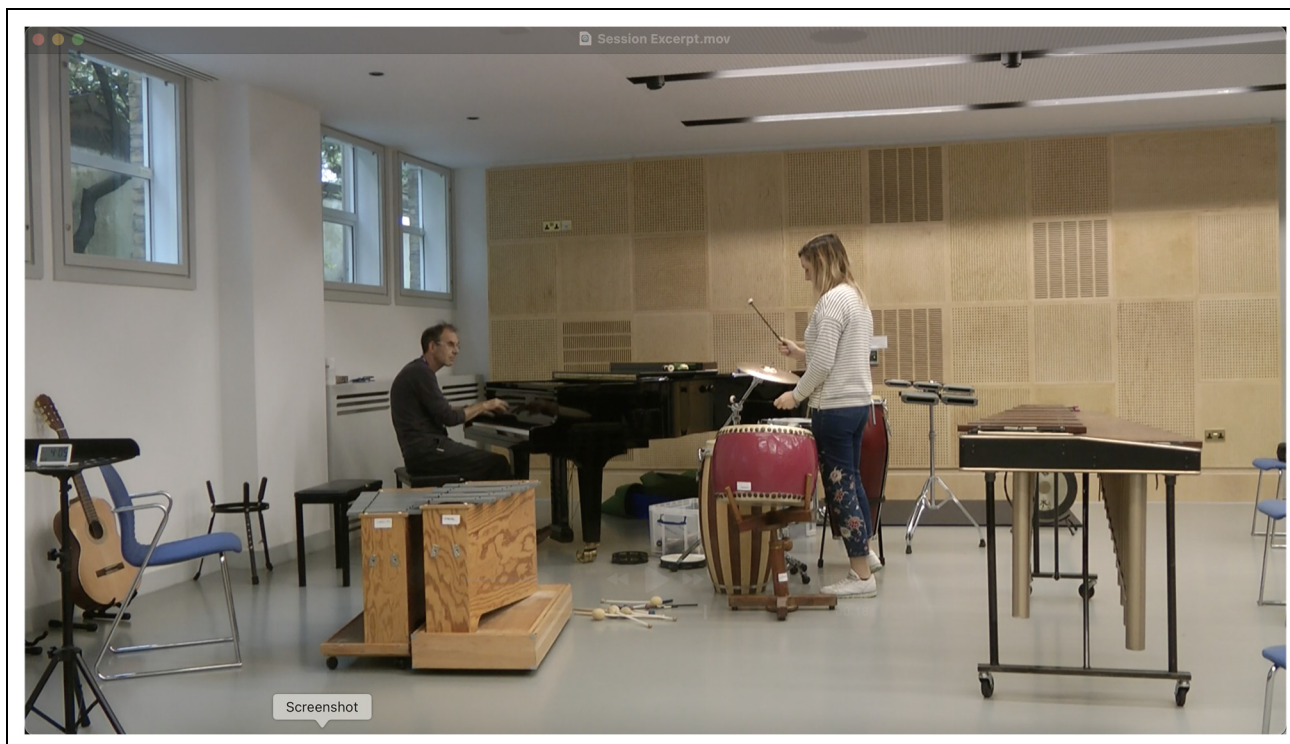


Figure 1. Still image from session excerpt (Video 1 in Supplemental Materials).

Analysis of videos of music therapy sessions—of full sessions and also of clips selected as demonstrating “pivotal moments” (Gavrielidou & Odell-Miller, 2017)—have been central to training and supervision of improvisatory music therapists in at least some schools of practice (e.g., Goodman, 2011).¹ Video clips not unlike this one, which was selected by an observing music therapist as noteworthy within a session, have also played an important role in communicating with members of the public, potential clients and their families, staff members, and funders of clinical practice about what happens in music therapy sessions and what kinds of transformations can be observed (e.g., Ledger, 2015, p. 855).² Clips tend to be selected for moments of what Bell et al. (2014) call “glaring salience,” even if moments in which less that is obvious is happening may also include important content (Bell et al., 2014). Observations of focus group reactions to video clips by family members of clients and non-music-therapy-trained professionals have suggested that what the music therapists thought was noteworthy and worth communicating about can indeed be readily apparent (e.g., Pavlicevic et al., 2014).

Our aim in the study presented here is to better understand how participants in and observers of music therapy sessions—both therapists and non-therapists—interpret what happened in the same way, and where their perspectives differ. We use as our material video generated for the current study with therapists trained in one form of music therapy (of the many that exist, see Bruscia, 2014;

McFerran et al., 2023; Spiro et al., 2014) characterized as Creative Music Therapy, a practice that is particularly “music-centered” and improvisational (Aigen, 2005; see also Hadar & Aigen, 2025; Pavlicevic, 1995, among many others). For ethical reasons—so as not to interfere with current clients’ therapy processes—we are careful to select non-therapist participants who were interested in music therapy but who had not been referred to or referred themselves to music therapy for any of the many reasons that people are referred, for example to address communication problems, depression, anxiety, or grief.

We approach this building on methods developed to uncover discrepancies and disagreements in how collaborating musicians in different genres understand what happened when they were making music together, and the ways in which outside observers overlap and do not overlap with the performers in their thinking about what happened (Pras et al., 2017; Schober & Spiro, 2014, 2016; Spiro & Schober, 2021). We also build on methods in music therapy research that incorporate multiple perspectives on specific moments in music therapy sessions (for example, perspectives of the participants and their families, researchers, and therapists), interrogating convergence and divergence to develop more nuanced understanding of what occurs (see, e.g., Flower, 2014; Rickson, 2022, among many others). In this study, we ask music therapists and session attendees to watch videos of a session they had just participated in to select and comment about moments they found notable, and

we ask non-session-participants (music therapists and non-music-therapy-trained observers) to also watch the videos, select moments, and comment. We later ask all study participants to rate the extent to which they endorse the full set of anonymized statements about the moments that all participants had selected, and to elaborate on their thinking in any cases where they do not endorse a statement.

The extreme glaring salience hypothesis is that everyone will see the same thing: that exactly what a therapist finds notable and selects as an example for untrained others will simply be obvious to everyone. While this is logically possible, we see this as unlikely always to be the case for several reasons. First, music therapists are advised to consult with other music therapists to see whether they identify the same episodes (Aldridge & Aldridge, 2008), and several hours of training can be needed even for music therapists to end up seeing sessions or particular moments in sessions more similarly (Bell et al., 2014; Turry et al., 2026). Second, playing together in music therapy is a very specific genre of music-making, one in which, it has been argued, actions by the therapists reflect clinical intentions beyond or different from those that a collaborating performer without clinical intentions—for example having the goal of making music for an audience—may have (Brown & Pavlicevic, 1996). From this perspective, therapists who understand music therapeutic clinical intentions should be more likely to see the same moments as of interest and to agree with each other's takes on what happened, and there should be less likelihood that non-therapist session attendees or observers should see things the same way.

In fact, given that clients don't necessarily have expectations for a therapy session that align with what therapists see as clinically useful options, one might expect that a client's interpretations should be especially different from the therapist's. This would be consistent with the claim that access to the interactive dynamics of an improvisation may be "more exclusively within the therapist's realm" (Pavlicevic, 1995, p. 172). In non-therapeutic contexts, there is evidence that interpretations of what happened by co-performing musicians can be far less aligned than musicians typically assume, and that players can even agree with an outsider's (a non-performing observer's)³ perspective on what happened more than with their collaborating partner's (Pras et al., 2017; Schober & Spiro, 2014; Spiro & Schober, 2021). With their different goals, music therapists and clients may think even more differently than collaborating musicians who share performance goals do, and they may agree with outsiders' perspectives even more.

Music therapy sessions, however, typically include more than just music-making: Therapists and clients can also interact with words, gestures, gaze, facial expressions, movements, and bodily alignments. Studies of spoken-language conversation have demonstrated how intertwined the linguistic and nonverbal behaviors of interlocutors are (Clark, 1996; Goodwin, 1986; Schegloff, 1987, among

many others). In terms of shared understanding in dialogue, participants in a conversation (addressees) with the goal of attempting to match figures can understand references reliably more accurately than overhearers who are not part of the dialogue, even if the overhearers have heard every word the participants have said since they met each other (Schober & Clark, 1989). To the extent that a music therapist and client form a dyad that shares experience and builds unique common ground, the interpretive dynamics of collaborative dialogue (Clark, 1996; Pickering & Garrod, 2004; Schober & Brennan, 2003) may apply to the music therapy situation. The therapist and client may share understanding that outsiders to a session—whether trained as music therapists or not—do not.

Based on prior evidence, we can also predict the converse: Listeners who are *not* participants in a musical interaction may be more likely to share interpretation of what happened than do the participants. In a study of 239 people who listened to audio recordings of three improvisations on a jazz standard by professional musicians working together for the first time (Schober & Spiro, 2016), listeners' patterns of judgment about statements authored by the performers and a musically trained outsider corresponded most with the outsider's judgments. If that were to be the case in the current setting, we would expect that the non-session-participants (the outsiders) should think more similarly about what happened in the sessions than the session-participants (music therapist and client) do.

Study

The dataset for this study consists of responses to a questionnaire about four videorecorded music therapy improvisation sessions, each with a different session music therapist who had undergone Nordoff–Robbins training and a session attendee who was not a trained music therapist and who had not participated in or been referred to music therapy before. We then invited another four music therapists, also with Nordoff–Robbins training, and four more people not trained as music therapists (one for each session) to be part of the study as commenters. We asked all 16 study participants to select moments from their session video to comment about, and then asked them to rate their endorsement of anonymized statements made by themselves and others about their session. This allows us to explore the following questions:

1. Do the study participants trained in music therapy (the session and commenting music therapists) think more like each other than the study participants who are not?
2. Do those who participated in a session (the music therapist and attendee) think more like each other than those who did not participate?

3. Conversely, do study participants who were *not* in the session (the commenting music therapist and the commenting observer) think more like each other?

Our setup allows other points of comparison, but these three questions hone in on potential effects of music therapy training and of participation. On a *music therapy training leads to thinking similarly* hypothesis, commenting music therapists who observe a videotape of a session should interpret what happened more similarly to the session music therapists than the untrained session attendees do. (A pattern like this would be consistent with evidence that topic expertise in dialogue—for example, describing New York landmarks for other New Yorkers vs. non-New-Yorkers [Isaacs & Clark, 1987]—can lead to improved and more efficient understanding through community-specific word choice). On a *session participation leads to thinking similarly* hypothesis, untrained session attendees should interpret what happened more similarly to the session music therapist than non-participating outsiders, just as addressees in dialogue can interpret references more accurately than overhearers (Schober & Clark, 1989). On a *listeners as outsiders* hypothesis (Schober & Spiro, 2016), listeners who were not involved in the session should interpret what happened more similarly than those who participated.

More complex patterns are also possible: If both music therapy training and session participation lead to greater similarity in interpretation, then untrained commenting outsiders should interpret what happened less like the music therapists than either the other music therapists or the untrained session attendees. And, of course, on an extreme glaring salience hypothesis, everyone should interpret what happened in the same way.

Our data set allows us to investigate similarity in interpretation using three measures:

1. Overlap in moments selected as worthy of commentary
2. Average endorsement ratings (on a five-point scale) of anonymized statements authored by study participants in the different roles
3. Level of agreement in the overall pattern of ratings of all the anonymized statements

It is important to note that we come to this research not as music therapists ourselves. We have consulted with practicing music therapists throughout the study—in developing the method and research questions, in recruiting music therapist participants, and in interpreting the findings—but recognize that we are likely to be missing nuances that practitioners would find. Also, it is not our intention to evaluate the quality or effectiveness of the sessions, which are of course different from actual first-time music therapy sessions that attendees might themselves have sought out or that might have been prescribed to address particular

concerns, and which might have been preceded by goal-setting conversations with the music therapists. That said, our perspective as non-music-therapists with interest in overlapping cognition in music-making is part of what allows us to ask these questions in the way we do. Indeed, we see this project as an example of how research can contribute to understanding what happens in music-making both in and beyond music therapy, from formal to informal settings, from individual lessons to community choirs to dance clubs to orchestra audiences—in the full range of what can be conceived as musical care practices (Spiro et al., 2023, in press; Spiro & Sanfilippo, 2022).

Method

Recruitment. Four music therapists who were working at the Nordoff–Robbins Music Therapy Centre in London, UK, at the time of data collection were invited, based on availability, to run a music therapy improvisation session with someone who had not previously experienced music therapy, as part of a research project. (See Figure 2 for the wording at the start of the information sheet they were sent during recruitment and that they again saw during the consenting process at the session.) The inclusion criteria were that they had completed music therapy training at Nordoff–Robbins, had been working as music therapists for at least two years, and would be available to run a session for this study in the Nordoff–Robbins London Centre.

The four first-time attendees were invited, through the authors' professional networks, to participate in a music therapy improvisation session (see Figure 2). The inclusion criteria were that they were interested in participating in a music therapy session, that they had not previously participated in music therapy sessions or trained as music therapists, that they had not been formally referred for music therapy, and that they would be able to attend a session in person for this study at the Centre. The music therapists and first-time attendees were offered £45 for their participation.

Four more music therapists also trained at Nordoff–Robbins and four more non-music-therapists were recruited to comment on a video of one of the study's music therapy improvisation sessions. (See Figure 3 for the wording at the start of the information sheet they were sent during recruitment and that they again saw during the consenting process at the session.) The inclusion criteria were the same in terms of music therapy experience, but this set of non-session-participants did not need to go to the Centre for their part in the study. The commenters were offered £30 for their participation.

The 16 study participants were assigned to four sessions (two music therapists and two non-music-therapists for each) according to availability.

Study participants. The eight music therapist participants had undertaken a two-year music therapy training course at Nordoff–Robbins and had all been working as music

Information Sheet Session Participants <i>The Purpose:</i> <i>This experiment aims to investigate how people think about music therapy sessions.</i> <i>Why have I been chosen?</i> You are being invited to take part in this research because we feel your experience can contribute to our understanding music therapy. <i>What do I have to do?</i> Your involvement in this study will be in three parts. 1) participate in a 30 minute session with a music therapist, which will be video and audio recorded, as either the music therapist or the attendee 2) annotate 1-5 important moments within the session and describe the reasons why they were chosen and 3) complete an individual online post-questionnaire based other participants' descriptions of the important moments. We may contact you after each step in order to make sure we have understood your answers.
--

Figure 2. Start of information sheet provided to the session music therapists and first time attendees.

Information Sheet Commenting participants <i>The Purpose</i> <i>This experiment aims to investigate how people think about music therapy sessions.</i> <i>Why have I been chosen?</i> You are being invited to take part in this research because we feel your experience can contribute to our understanding music therapy. <i>What do I have to do?</i> Your involvement in this study will be in three parts. 1) Observation of a video and audio recording of a 30-minute session with a music therapist 2) annotation of 1-5 important moments within the session and describe the reasons why they were chosen and 3) completion of an individual online post-questionnaire based other participants' descriptions of the important moments. We may contact you after each step in order to make sure we have understood your answers.
--

Figure 3. Start of information sheet provided to all commenters.

therapists for at least two years; one music therapist had also undertaken additional music therapy training elsewhere. The music therapists had worked together at Nordoff–Robbins but not directly in a therapy room together. The music therapists and attendees had never played together before; three of the four attendees had never met the music therapist in their session, and the one who had did not know them well.

As Table 1 details, all 16 study participants reported (on post-session questionnaires) having played and currently playing at least one instrument regularly, and all reported having played between 1 and 4 h daily during their peak playing period. The music therapists had all been playing for over 10 years and reported experience improvising musically. The first-time attendees and non-attending observers ranged in years of musical experience and formal musical training; four reported little or no experience improvising,

and four reported “a lot” or “a great deal.” The music therapists had experience playing a greater range of instruments. See Table 2 for additional self-reported demographic characteristics of the sample of study participants.

Procedure

Phase 1: Session and Post-Session Characterizations. *Music therapists and session attendees.* The sessions, which took place in October and November 2017 in dedicated music therapy rooms at Nordoff–Robbins, London, UK, were scheduled for 30 min, each one with a different music therapist and a different first-time attendee. Before each session, the music therapist came early to give their consent to participate and to set up the room. The music therapists set up the room as they would for a typical first session with an

Table 1. Self-Reported Music Therapy and Musical Experience.

	Music Therapists (<i>n</i> = 4)	Attendees (<i>n</i> = 4)	Commenting Music Therapists (<i>n</i> = 4)	Commenting Observers (<i>n</i> = 4)
Music therapy training				
2–10 years working as a music therapist	2	-	3	-
10 or more years working as a music therapist	2	-	1	-
Music therapy participation experience				
Previously participated in a music therapy session	4		4	
Previously watched music therapy session	4	1	4	2
Like music				
Like a great deal	4	4	3	4
Instruments regularly play now				
Voice	4	3	3	3
Guitar (acoustic or electric)	4	2	3	2
Piano	3		3	3
Ukulele	2		1	
Electronic Keyboard	2		2	
Bass guitar	1	1	2	
Drums	1		1	1
Xylophone	1		1	
Organ	1			
Clarinet	1			
Violin		1		1
Banjo			1	
Bazouki			1	
Tin whistle		1		
Number of instruments regularly play now (mean, SD)	5, 1.15	2, 0.82	4.5, 2.65	2.5, 1
Instruments have ever played				
Voice	2	2	3	3
Guitar (acoustic or electric)	3	2	2	2
Piano	3		2	3
Bass guitar	1	2	2	
Electronic Keyboard	1	1	1	
Ukulele	2		1	
Violin		1		1
Drums	1		1	1
Saxophone	2			
Clarinet	2		2	
Organ	1			
Xylophone	1			
Banjo			1	
Bazouki			1	
Oboe	1			
Tin whistle				
Average number of instruments ever played (mean, SD)	5, 3.16	2.75, 1.5	3.75, 2.06	2.5, 1
Years of regular daily playing				
0 years				
Up to 2 years				1
2 to 5 years		1		
5 to 10 years		3		
10 or more years	4		4	3
Number of hours playing per day during peak playing period				
1 to 2 h	2	1	2	2
2 to 4 h	2	3	2	2
Experience improvising musically				
None		1		1
A little		2		
Some			1	
A lot	2	1	1	1

(continued)

Table 1. Continued.

	Music Therapists (n = 4)	Attendees (n = 4)	Commenting Music Therapists (n = 4)	Commenting Observers (n = 4)
A great deal	2		2	2
Experience playing in duos				
None				1
A little		1		
Some	1	3	1	
A lot	1			
A great deal	2		3	3
Experience playing in ensembles				
None				1
A little				
Some	1	1	1	1
A lot	2	3	1	1
A great deal	1		2	1
Experience playing in jam sessions				
None	1		1	1
A little		3		
Some			1	1
A lot	3	1		
A great deal			2	3
Years formal training on a music instrument				
0 years			1	1
Up to 2 years		1		1
2 to 5 years				
5 to 10 years	1	2	2	
10 or more years	3	1	1	2
Years of formal training in music theory				
0 years		1	1	
Up to 2 years				2
2 to 5 years	2	3		
5 to 10 years	1		3	2
10 or more years	1			
Write about music				
Rarely		1	1	
Sometimes	2	1		2
Frequently	1		1	2
Regularly	1	2	2	
Genres of music play				
Rock/pop	4	2	2	1
Jazz	3	1	2	2
Classical	3	1	2	3
Folk	2	3	3	1
Soul, Ambient		1		
Hip-Hop Funk			1	
Genres of music choose to listen to				
Rock/pop	4	3	2	3
Jazz	3	1	2	2
Classical	3	2	1	2
Folk	4	2	3	1
Hip Hop			1	1
Electronic, alternative, Nu soul		1		
Musicals		1		
R&B, soul, acoustic		1		

Table 2. Self-Reported Demographic Characteristics.

	Music Therapists (n = 4)	Attendees (n = 4)	Commenting Music Therapists (n = 4)	Commenting Observers (n = 4)
Gender				
Female	2	3	2	3
Male	2	1	2	1
Race/ethnicity				
Asian/Pacific Islander				1
Caucasian/White	4	3	4	3
Hispanic		1		
Education				
Undergraduate / Bachelors Degree (including performance)		1		
Masters degree (including performance)	4	3	4	4
Comprehension of English language (1 = Poor, 5 = Excellent)				
Scoring 5 (Excellent)	4	4	4	4

**Figure 4.** Photograph of the setup for one of the four sessions.

adult (see Figure 4 for an example). As in many music therapy sessions, there was a wide range of tuned and untuned instruments in the room: a piano, guitars, xylophones, shakers, drums, and other percussion instruments. A video camera and external microphone for recording the session were in the room and visible to the session participants; this is not unusual in actual music therapy sessions where an attendee's consent for recording is given.

Session attendees arrived separately and provided their consent to participate. Just before the session, they were

brought to the music therapy room and introduced to the music therapist. The researchers left the room and the session then commenced. When the music therapist ended the session, the researchers escorted the music therapist and the session attendee to different rooms for their independent review of the session.

Post-Session Characterizations: Session Participants. Immediately after the session was over, the session participants were taken to different, quiet rooms by one of the researchers

surveymonkey.co.uk/r/importantmoments

What moments in the session stick out for you for whatever reason? Please indicate a time period during which something of note happens. Please be as precise as possible with the times. Choose 1-5 different moments and write down when they start and finish (using the time on Quicktime), and why you have chosen that moment. Feel free to discuss anything that stuck out to you. For example you can comment on the music, on how those participating felt, on the people, a combination of these or anything else that struck you. (Feel free to fill this out in whatever way is easier for you.)

Moment 1: Start time

Moment 1: End time

Reason I chose moment 1

Figure 5. Excerpt from one Phase 1 questionnaire for selecting and describing 1 to 5 important moments in the session.

and had a short break while the videos were transferred from the camera's memory stick to a laptop provided for each session participant. Session participants were then provided with headphones and left alone to fill in an online questionnaire in the online survey creation platform Survey Monkey. The questionnaire asked participants to watch and listen to as much or as little of the video (visible in a separate window) as they wanted in order to identify between one and five moments that struck them for whatever reason (Figure 5); they could comment on the music, on how those participating had felt, on the people or a combination of these or anything else.⁴ They were asked to provide the exact start and end time of each moment based on the time stamps in the video. They were told that there was no set time limit, that they could take breaks, and that this was not a memory test; they could stop, start, and replay as much as they liked.

The questionnaire ended with an open question for any further comments about the session as well as questions that verified participants' music therapy training and how well they knew the other session participant. It took between 1 and 2 h for session participants to complete the survey.

Before they departed, the session participants were reminded not to discuss the research project with their session partner. The music therapists were also reminded not to discuss the research with any other music therapists until after the project was complete.

Post-Session Characterizations: Non-Session-Participant Commenters. After each session was over (either on the same or the next day), the two commenters per session were emailed a password-protected link to the session video along with a link to the same online questionnaire that the session participants had completed. They were asked to complete the questionnaire in a quiet room, with headphones if that would be helpful. They received the same instructions as the session participants, including about being able to rewatch as much as they liked and being asked to provide precise start and end times for the moments they selected.

Non-session-participants were asked to complete the task as soon as possible. All non-session-participants did so within six weeks of the music therapy session they were commenting on.

Phase 2: Rating Anonymized Characterizations. In May and June 2018, all study participants were emailed a link to a second questionnaire, implemented in the online survey creation platform Qualtrics, developed from anonymized versions of the comments made by all four commenters for their session. Each questionnaire (one for each of the four sessions) included the video excerpts that corresponded with the moments that all study participants connected with a session had commented on, and each question was connected in the survey software with video of the moment that it was about. Because a number of comments about

single moments were multi-statement and longer than anticipated, the questionnaire presented individual component statements to be rated separately. To create the component statements, we followed these guidelines:

1. Use study participant's wording as much as possible, but do not include wording that would identify the author or the author's role (therapist, session attendee, observer).
2. Label session participant roles consistently: "music therapist" or "session participant," or "players" when referring to both.⁵ Replace all names and personal pronouns (I, me, us) with role labels.
3. Standardize spelling (correct typos) and verb tense (all present tense).
4. Aim for each statement to be ratable on its own. If necessary, include (repeat) wording from a prior statement to disambiguate.
5. Omit words of introduction that do not add ratable content (e.g., "I felt this moment was significant because...").
6. Do not include statements that make general claims about music therapy beyond the specific moments (e.g., "The process which the players have gone through is the cornerstone of the Nordoff–Robbins approach").

This process led to questionnaires with 139 ratable statements (Session 1), 123 statements (Session 2), 106 statements (Session 3), and 156 statements (Session 4).⁶

In the questionnaires, for most statements the rating task was introduced with this prompt: "To what extent do you agree with the following characterizations of this moment ([start time – end time])?". Study participants were asked to rate agreement on a five-point scale (*Strongly disagree*, *Disagree*, *Neither agree nor disagree*, *Agree*, *Strongly agree*, with two additional options: *Don't understand the statement* and *Can't judge*). (See Figure 4). For each moment that included statements judged by the authors as being about thoughts and feelings, a second questionnaire screen was presented that began with "Even if it's not possible to know for sure what someone else is thinking or feeling, to what extent do you think the following characterizations of this moment ([start time – end time]) are likely to be accurate?" For these questions, study participants were again asked to rate agreement on a five-point scale with labels *Strongly unlikely*, *Unlikely*, *Neither likely nor unlikely*, *Likely*, *Strongly likely*, and with the same two additional options *Don't understand the statement* and *Can't judge*). (See Figure 6).

In cases where moments selected by more than one commenter overlapped in time, the video excerpt in the

To what extent do you agree with the following characterisations of this moment (10:25-10:43)?

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Don't understand the statement	Can't judge
The session participant has more musical possibility than before as she has a range of tuned and untuned percussion around her and 2 beaters, one in each hand	☐	☐	☐	☐	☐	☐	☐
This piece of joint music making gets into a flow quickly	☐	☐	☐	☐	☐	☐	☐
The session participant explores the melodic and rhythmic possibilities of her instruments	☐	☐	☐	☐	☐	☐	☐
Both players find a mutual ending	☐	☐	☐	☐	☐	☐	☐
The music therapist initiates the mutual ending with a tiny slowing down	☐	☐	☐	☐	☐	☐	☐
The session participant picks up the therapist's slowing down	☐	☐	☐	☐	☐	☐	☐
The players look at one another	☐	☐	☐	☐	☐	☐	☐
The players slow down to a mutually placed ending together	☐	☐	☐	☐	☐	☐	☐

Even if it's not possible to know for sure what someone else is thinking or feeling, to what extent do you think the following characterisations of this moment (10:25-10:43) are likely to be accurate?

	Strongly unlikely	Unlikely	Neither likely nor unlikely	Likely	Strongly likely	Don't understand the statement	Can't judge
The session participant feels much freer in her music making now	☐	☐	☐	☐	☐	☐	☐
The session participant hears the therapist's slowing down	☐	☐	☐	☐	☐	☐	☐
The players listen carefully	☐	☐	☐	☐	☐	☐	☐

Figure 6. Excerpt from one Phase 2 questionnaire that shows how the video and the two kinds of rating questions were displayed. When study participants clicked the "Play" arrow, the video (same material as in Video 1 in this case) started at the timestamp listed; study participants could rewatch or watch preceding and subsequent context in the video if they chose.

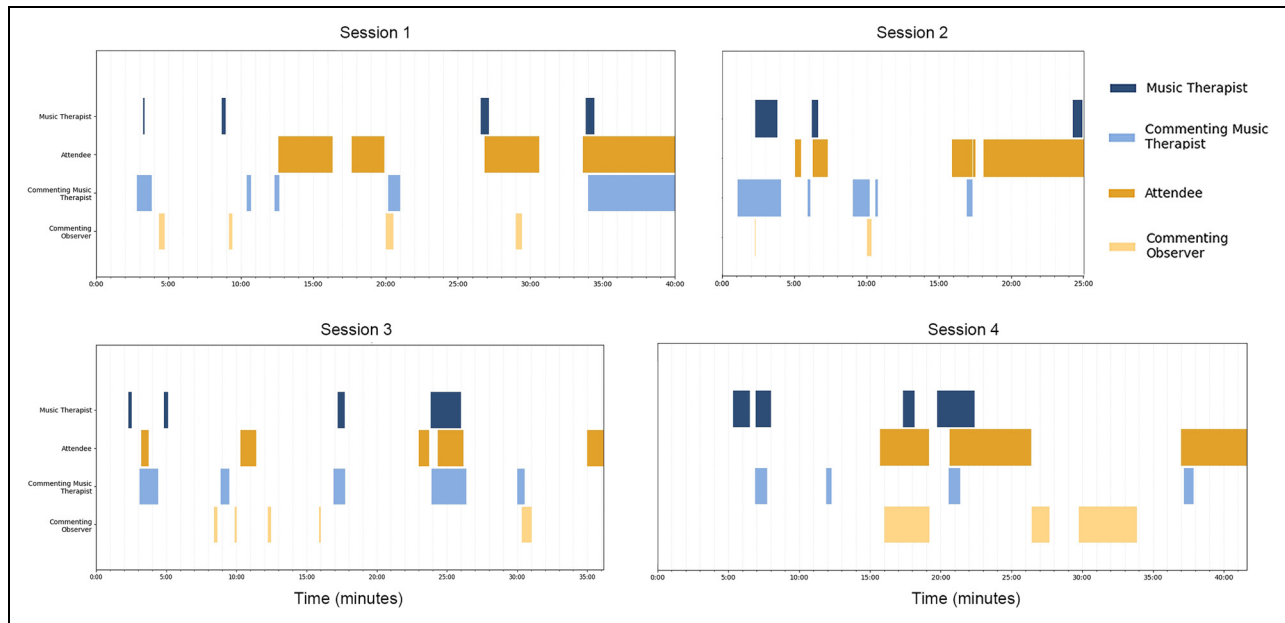


Figure 7. Start and end times for each moment selected by each study participant. the music therapists' choices of moments are shaded blue, and the non-music-therapists' moments are shaded orange.

questionnaire started at the time stamp where the first comment began and ended at the time stamp where the last moment ended.

Figure 6 gives one example of how original long comments were broken into individual statements suitable for rating when that was needed. Seven of the statements (the final 5 of the first set and the final 2 of the second set) are derived from the original long comment "I felt this moment was significant as they both find a mutual ending, which is initiated by a tiny slowing down in the therapists playing, that is heard and picked up by the client who slows down too, they are looking at one another and listening carefully and slow down to a mutually placed ending together."

Following all the ratings, study participants were asked to elaborate on why they had disagreed (selected *Strongly disagree* or *Disagree*, or *Strongly unlikely* or *Unlikely*) with the prompt "Looking back at moment [start time – end time], for the following statement: [statement repeated] You selected: [answer selected] Can you tell us why you chose this answer?" As before, the video was presented timed to start at the start time of the moment referred to. The requests for elaboration were presented after all the ratings so as not to disincentivize disagreement in the ratings section.

The final section of the questionnaire asked about participants' demographic characteristics, their music therapy experience and their musical experience as well as about their experience of the session and of answering this questionnaire.

Results

1. Overlap in moments selected as worthy of commentary

Figure 7 maps the start and end times for each moment selected by each study participant. In all four sessions, no one selected exactly the same moments to comment on. But there were some parts of each session that attracted more commentary than others: In each group, there were one or two moments that three of the four study participants commented on.

Tables 3 and 4 quantify the degree of overlap in moments selected by the different study participants by counting two moments as overlapping if they include any of the same seconds at all (so even a one-second overlap counts as an overlap). As seen in the tables (Table 3 presents average percentages across the four sessions, and Table 4 presents the findings for each session individually), the data show modest support for a *music therapy training leads to thinking similarly* hypothesis. Considering the moments that the session music therapist selected (either 4 or 5 moments in a session), the commenting music therapists' moment selections overlapped on an average of 43% of them, ranging from 33% to 50%. Considering the moments that the commenting music therapist selected, the session music therapist's moment selections overlapped on an average of 38% of them, ranging from 20% to 50%. This degree of overlap is substantially higher than either music therapist's generally low level of overlap with the commenting outsider's moments selected.

Table 3. Average (Across the Four Sessions) in the Number of Moments Selected by Each Study Participant that Overlap for at Least One Second with Moments Selected by Each Other Participant.

Overlap with Moments Selected by...	Music Therapist	Attendee	Commenting Music Therapist	Commenting Observer
Music therapist		47%	43%	15%
Attendee	44%		44%	15%
Commenting music therapist	38%	38%		20%
Commenting observer	0%	15%	36%	

Table 4. Percentage and Number of Moments Selected by Each Study Participant That Overlap for One Second or More with Moments Selected by Each Other Participant.

Overlap with Moments Selected by...	Music Therapist	Attendee	Commenting Music Therapist	Commenting Observer
Session 1				
Music therapist (<i>n</i> = 4)		50% (<i>n</i> = 2)	50% (<i>n</i> = 2)	0%
Attendee (<i>n</i> = 4)	50% (<i>n</i> = 2)		50% (<i>n</i> = 2)	25% (<i>n</i> = 1)
Commenting music therapist (<i>n</i> = 5)	40% (<i>n</i> = 2)	40% (<i>n</i> = 2)		20% (<i>n</i> = 1)
Commenting observer (<i>n</i> = 4)	0%	25% (<i>n</i> = 1)	25% (<i>n</i> = 1)	
Session 2				
Music therapist (<i>n</i> = 3)		67% (<i>n</i> = 2)	33% (<i>n</i> = 1)	33% (<i>n</i> = 1)
Attendee (<i>n</i> = 5)	40% (<i>n</i> = 2)		20% (<i>n</i> = 1)	0%
Commenting music therapist (<i>n</i> = 5)	20% (<i>n</i> = 1)	20% (<i>n</i> = 1)		40% (<i>n</i> = 2)
Commenting observer (<i>n</i> = 2)	50% (<i>n</i> = 1)	0%	100% (<i>n</i> = 2)	
Session 3				
Music therapist (<i>n</i> = 5)		20% (<i>n</i> = 1)	40% (<i>n</i> = 2)	0%
Attendee (<i>n</i> = 5)	20% (<i>n</i> = 1)		40% (<i>n</i> = 2)	0%
Commenting music therapist (<i>n</i> = 5)	40% (<i>n</i> = 2)	40% (<i>n</i> = 2)		20% (<i>n</i> = 1)
Commenting observer (<i>n</i> = 5)	0%	0%	20% (<i>n</i> = 1)	
Session 4				
Music therapist (<i>n</i> = 4)		50% (<i>n</i> = 2)	50% (<i>n</i> = 2)	25% (<i>n</i> = 1)
Attendee (<i>n</i> = 3)	67% (<i>n</i> = 2)		67% (<i>n</i> = 2)	33% (<i>n</i> = 1)
Commenting music therapist (<i>n</i> = 4)	50% (<i>n</i> = 2)	50% (<i>n</i> = 2)		0%
Commenting observer (<i>n</i> = 3)	33% (<i>n</i> = 1)	33% (<i>n</i> = 1)	0%	

The overall pattern also shows modest support for the *session participation leads to thinking similarly* hypothesis in that the session music therapists' selected moments overlapped with an average of 44% of the attendees' selected moments, ranging from 20% to 67%, and the attendees' selected moments overlapped with an average of 47% of the session music therapist's selected moments, again ranging from 20% to 67%. These averages are in the same range as those between the two music therapists, and notably higher than the averages with the commenting outsider.

We also see modest support for a pattern of session-outsider agreement (the *listeners as outsiders* hypothesis): The commenting music therapists overlapped, on average, with 36% of the moments selected by the commenting observer, which is much more than the music therapists' average of 0% and the attendees' average of 15%. Given how much variability there is across the four sessions, we are cautious about overinterpreting this, but this figure does stand out from the others.

Another way of looking at the overlap in moments selected is in the number of seconds that study participants selected as worthy of commentary that overlapped. Table 5 presents percentages of seconds in moments selected by each study participant that overlap with seconds in moments selected by each other participant, averaged across the four sessions, and Table 6 presents the number of seconds and percentage values for each session. From this perspective, we again see moderate support for the *training leads to thinking similarly* perspective. The commenting music therapists' selected moments included an average of 50% of the same seconds as the session music therapists', ranging from 28% to 80%, and the session music therapists' selected moments included an average of 36% of the same seconds as the commenting music therapists', ranging from 6% to 60%. These averages were in all sessions higher than the average level of overlap between either music therapist and the commenting outsider.

Table 5. Average (Across the Four Sessions) Percentage of Seconds in Moments Selected by Each Study Participant That Overlap with Seconds in Moments Selected by Each Other Participant.

Percentage Overlap of Seconds in Moments Selected by	Music Therapist	Attendee	Commenting Music Therapist	Commenting Observer
Music therapist		49%	50%	4%
Attendee	16%		24%	7%
Commenting music therapist	36%	44%		3%
Commenting observer	4%	16%	25%	

Table 6. Percentage and Number of Seconds in Moments Selected by Each Study Participant That Overlap with Seconds in Moments Selected by Each Other Participant.

Overlap of Seconds Moments Selected by	Music Therapist	Attendee	Commenting Music Therapist	Commenting Observer
Session 1				
Music therapist (90 s)		59% (53 s)	33% (30 s)	0% (0 s)
Attendee (997 s)	5% (53 s)		37% (365 s)	3% (26 s)
Commenting music therapist (508 s)	6% (30 s)	72% (365 s)		4% (22 s)
Commenting observer (95 s)	0% (0 s)	27% (26 s)	23% (22 s)	
Session 2				
Music therapist (158 s)		39% (62 s)	58% (92 s)	1% (1 s)
Attendee (613 s)	10% (62 s)		4% (23 s)	0% (0 s)
Commenting music therapist (293 s)	31% (92 s)	8% (23 s)		4% (13 s)
Commenting observer (21 s)	5% (1 s)	0% (0 s)	62% (13 s)	
Session 3				
Music therapist (194 s)		52% (100 s)	80% (156 s)	0% (0 s)
Attendee (324 s)	31% (100 s)		44% (142 s)	0% (0 s)
Commenting music therapist (347 s)	45% (156 s)	41% (142 s)		4% (13 s)
Commenting observer (88 s)	0% (0 s)	0% (0 s)	15% (13 s)	
Session 4				
Music therapist (344 s)		46% (157 s)	28% (97 s)	15% (50 s)
Attendee (833 s)	19% (157 s)		10% (85 s)	23% (189 s)
Commenting music therapist (162 s)	60% (97 s)	53% (85 s)		0% (0 s)
Commenting observer (513 s)	10% (97 s)	37% (189 s)	0% (0 s)	

We also again see some support for a *session participation leads to thinking similarly* hypothesis, but from this perspective the pattern is not symmetrical. On average the attendees overlapped in 49% of the seconds in the music therapists' selected moments, ranging from 39% to 59%, but the music therapists overlapped in only 16% of the seconds that the attendees selected, running from 5% to 31%. On average the attendees' moments overlapped in seconds with the commenting music therapists' moments at a rate closer to that with the session music therapists, 44%, ranging from 8% to 72%. So it seems that the attendees selected a greater percentage of the same moments as either music therapist, whether or not the therapist had participated in the session.

Both views of the moments selected show that commenting non-therapist observers' perspectives on what is worthy of commentary are particularly different from either the music therapists' or the attendees'. The patterns of overlap

are not consistent with a glaring salience hypothesis. The overall pattern suggests that training and participation both lead to some similarity in thinking, and that session-outsider status may also lead to some similarity in thinking. But the overlap is far from total in any of these pairings.

2. Endorsement ratings of session music therapists' statements

Table 7 shows the endorsement ratings of anonymized statements by each author averaged across the four sessions, and Table 8 shows each session's ratings. Ratings (1 for *Strongly disagree* or *Strongly unlikely* to 5 for *Strongly agree* and *Strongly likely*) averaged above neutral (3.0) and often above "agree" (4.0), but strong agreement with all or even many statements (5) was rare. Study participants

Table 7. Average Endorsement Ratings on 1–5 Scale of Anonymized Statements by Each Author, Excluding Ratings of “Don’t Understand the Statement” and “Can’t Judge.” Ratings of Statements Authored by the Rater are Enclosed in Parentheses.

Rater				
Author	Music Therapist	Attendee	Commenting Music Therapist	Commenting Observer
Music therapist	(4.43)	4.38	3.96	3.90
Attendee	4.13	(4.63)	4.17	4.05
Commenting music therapist	4.28	4.42	(4.58)	3.90
Commenting observer	3.79	4.21	3.64	(3.94)

Table 8. Average Endorsement Ratings on 1–5 Scale of Statements by Each Author per Session, Excluding Ratings of “Don’t Understand the Statement” and “Can’t Judge.” Ratings of Statements Authored by the Rater are Enclosed in Parentheses.

Rater				
Author	Music Therapist	Attendee	Commenting Music Therapist	Commenting Observer
Session 1				
Music therapist	(4.26)	4.14	4.09	3.76
Attendee	3.97	(4.66)	4.09	3.56
Commenting music therapist	4.25	4.41	(4.83)	3.86
Commenting observer	3.67	4.11	3.78	(4.25)
Session 2				
Music therapist	(4.43)	4.76	4.14	4.11
Attendee	4.10	(4.86)	4.19	4.06
Commenting music therapist	4.14	4.68	(4.68)	4.15
Commenting observer	4.33	4.00	5.00	(4.33)
Session 3				
Music therapist	(4.42)	3.91	3.47	3.76
Attendee	4.33	(4.00)	3.86	4.56
Commenting music therapist	4.49	3.95	(3.97)	4.06
Commenting observer	3.20	3.50	2.62	(3.47)
Session 4				
Music therapist	(4.58)	4.51	4.09	3.93
Attendee	4.56	(4.44)	4.75	4.00
Commenting music therapist	4.29	4.39	(4.61)	3.55
Commenting observer	4.20	4.41	4.05	(4.43)

mostly endorsed the anonymized versions of statements they themselves had authored (the diagonal values in the tables) more than they endorsed statements authored by others (higher mean agreement ratings), with a few exceptions. (Here we report complete data from all study participants but one; a technical problem led to missing data from one participant, the commenting observer in session 4, for the final 46 statement ratings. The complete data set consists of 1940 ratings across the four sessions).

Looking at the rows of the tables tells a story about which other participants most endorsed anonymized statements authored by participants in each role (with the caveat that we are looking at data generated by a small number of participants about four sessions). For statements authored by the session music therapists, on average and in all four sessions the attendee is the participant

who endorsed these statements most. For statements authored by the attendee, it is the commenting music therapists who give the highest endorsement ratings both on average and in three of the four sessions. In three of the four sessions, and on average, the attendee is the participant who endorses the commenting music therapist’s statements most. For statements authored by the commenting observer, it is the attendee who gives the highest endorsement ratings both on average and in three of the four sessions.

Looking at the columns of the tables gives another perspective: Which anonymized statements authored by other participants did participants in each role most endorse? The session music therapists on average most endorsed statements authored by the commenting music therapists, which was also the pattern in two of the four

Table 9. Cohen's Kappas Measuring Interrater Agreement for Each Participant on All the Statements They Both Rated (n). Kappas Significantly Different from 0 are Bolded.

	Music Therapist – Attendee	Music Therapist – Commenting Music Therapist	Music Therapist – Commenting Observer	Attendee – Commenting Music Therapist	Attendee – Commenting Observer	Commenting Music Therapist – Commenting Observer
Session 1	0.014 $p = 1.000$ $n = 139$	0.216 $p = .000$ $n = 139$	0.124 $p = .010$ $n = 139$	0.124 $p = .063$ $n = 139$	0.066 $p = .023$ $n = 139$	0.160 $p = .000$ $n = 139$
Session 2	–0.055 $p = .524$ $n = 123$	0.117 $p = .094$ $n = 123$	0.108 $p = .026$ $n = 123$	–0.039 $p = .584$ $n = 123$	0.198 $p = .005$ $n = 123$	0.146 $p = .005$ $n = 123$
Session 3	0.082 $p = .007$ $n = 106$	0.193 $p = .001$ $n = 106$	0.132 $p = .047$ $n = 106$	0.151 $p = .008$ $n = 106$	0.084 $p = .105$ $n = 106$	0.234 $p = .000$ $n = 106$
Session 4	0.130 $p = .019$ $n = 156$	0.213 $p = .002$ $n = 156$	0.060 $p = .289$ $n = 110$	0.175 $p = .001$ $n = 156$	0.064 $p = .194$ $n = 110$	0.131 $p = .020$ $n = 110$
Average	0.04275	0.18475	0.106	0.10275	0.103	0.16775

sessions. The session attendees most endorsed statements by one or the other of the music therapists, a pattern which held in three of the four sessions. The commenting music therapists most endorsed statements authored by the attendees, on average, and most in two of the sessions. (In one session it is a draw, and in another the commenting music therapist most endorses statements authored by the commenting observer). The commenting observers' ratings were generally lower and don't have a clear pattern in whose statements they endorse more.

To what extent are these patterns consistent with a music therapy *training leads to thinking similarly* hypothesis? Only partially. The music therapists most endorsed statements by the commenting music therapists, but the commenting music therapists did *not* most endorse statements by the session music therapists, and their ratings of session music therapists' statements were not higher than ratings by the other participants. And what about the *session participation leads to thinking similarly* hypothesis? Again, the data are only partially consistent. The untrained session attendees did interpret what happened more similarly to the session music therapists than did the non-participating outsiders, but the session music therapists did not endorse the session attendees' statements the most. So the thinking similarly among music therapists and among session participants (music therapists and attendees) is not symmetrical.

As we saw in the pattern of study participants' selecting the same moments as worthy of commentary, ratings of statements authored by the non-therapist commenting observer averaged the lowest endorsement ratings by participants in the other three roles. This was also the case in most individual sessions, though not uniformly.

Endorsement ratings by the non-therapist commenting observers were also, on average, lower than those of the other participants. In general, we see this pattern as consistent with a view that non-therapist commenting observers do not have the kind of session-insider access that music therapists or session attendees do. With this measure, we also do not see evidence that the outsiders think more similarly.

3. Agreement with others in their patterns of judgments

Another way of looking at similarity in thinking is in the extent to which study participants' overall patterns of ratings of all statements were similar. A straightforward way to examine this is to measure interrater agreement with Cohen's kappa, which ranges from –1 (complete disagreement) to +1 (complete agreement), and which assesses statistically significant agreement relative to the agreement that could occur by chance given the number of categories and ratings. If music therapists think more similarly, then their overall kappas with each other should be higher than kappas between their ratings and those of the other participants. If session participation leads to thinking similarly, then the session participants' kappas with the session music therapists should be higher than either of their kappas with the non-session-participants. If session outsiders think more similarly, then the commenting music therapists' and commenting non-therapist observers' kappas should be higher than the others.

Table 9 shows interrater agreement with kappas calculated following procedures used in previous studies of musical participants' levels of agreement across multiple

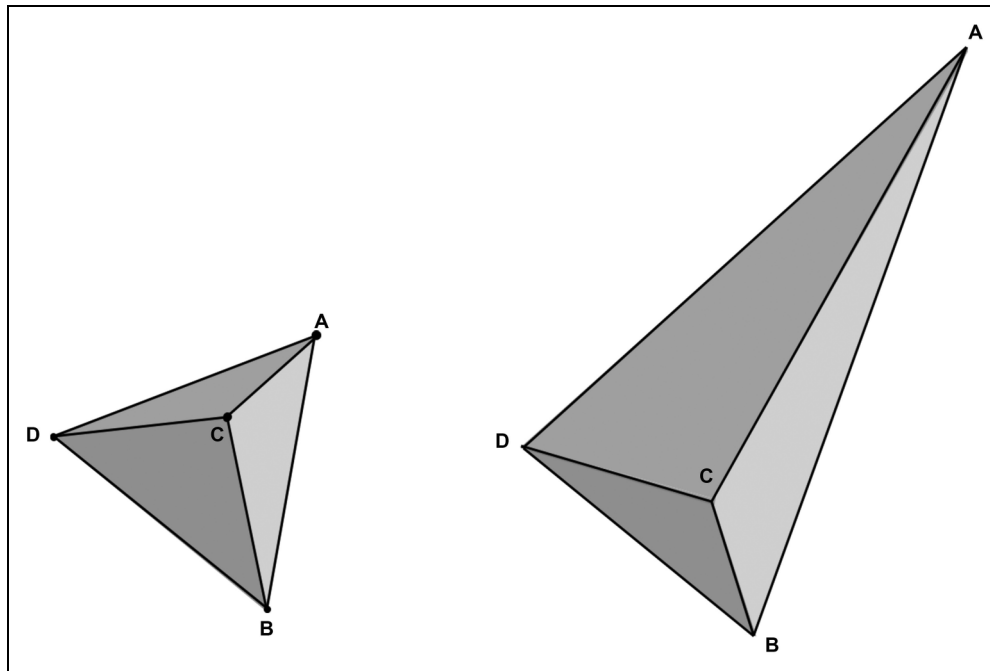


Figure 8. (a) Hypothetical tetrahedron representing equal kappas (equidistant edges). (b) Hypothetical tetrahedron representing one party (A) with low agreement with the other three participants, whose kappas with each other are equal.

comments about a performance (Pras et al., 2017; Schober & Spiro, 2014); two ratings are counted as agreeing if both participants disagreed with a statement (rating 1 or 2), both neither agreed nor disagreed (3), both agreed (rating 4 or 5), or both found the statement uninterpretable (rating 6 or 7).⁷ Here we again report complete data from all study participants but with fewer data points for session 4, because of the technical problem that led to some missing data from one participant.

What is immediately apparent is that two thirds of the kappa values (16 of 24) are significantly greater than chance. It is also apparent that by standard interpretations (Fleiss, 2003; Landis & Koch, 1977) the kappas are quite low, indicating only “slight” agreement, with only two of the 24 reaching the level of “fair” (above .21 in both Landis & Koch’s and Fleiss’s schemes). These levels of agreement are lower than those observed in statements made about jazz and free jazz improvisations by professional performers (Pras et al., 2017; Schober & Spiro, 2014).

Looking at the average kappas across the four sessions— noting that the ranking of kappas differs in each session—the highest agreement is between the two music therapists, reaching “fair” agreement in two of the four sessions. Close behind is agreement between the two session outsiders—the commenting music therapist and the commenting non-therapist observer—in their pattern of ratings, all significantly greater than chance agreement and reaching “fair” agreement in one session. Notably lower are the average levels of agreement between the session music therapists and commenting non-therapist observers, the attendees

and the commenting music therapists, and the attendees and commenting non-therapist observers. Lowest of all are the levels of agreement between the session music therapists and the attendees.

To visualize the interrelationships of each party with each other simultaneously, we can plot the extent of interrater agreement between any two parties by transforming the kappas into distances, with higher kappas represented through shorter distances between nodes on a graph, and lower kappas (less agreement) represented through greater distances. With four participants in a session, we can thus represent each participant’s relative agreement with the other as a tetrahedron. Because of the constraint that the transformed agreement values must occur within the bounds of what is physically possible in a tetrahedron, we simplify the relationships with stylized tetrahedra that map the distances onto a simpler scheme that fits physical constraints, representing the highest kappas (greatest agreement, kappas above .16) with a distance of 2, medium kappas (.09 to .159) with a distance of 3, and the lowest kappas (below .09) with a distance of 4. With this simplification, the tetrahedron shape that would occur if all four participants were to agree with each other equally would be a tetrahedron with fully equal sides (Figure 8a). If three participants were close in judgment with each other and far in judgment from the remaining participant, the tetrahedron shape we should observe would look more like a very tall pyramid (Figure 8b).

So what shaped tetrahedra do we observe? Figure 9 shows the tetrahedron based on the average levels of

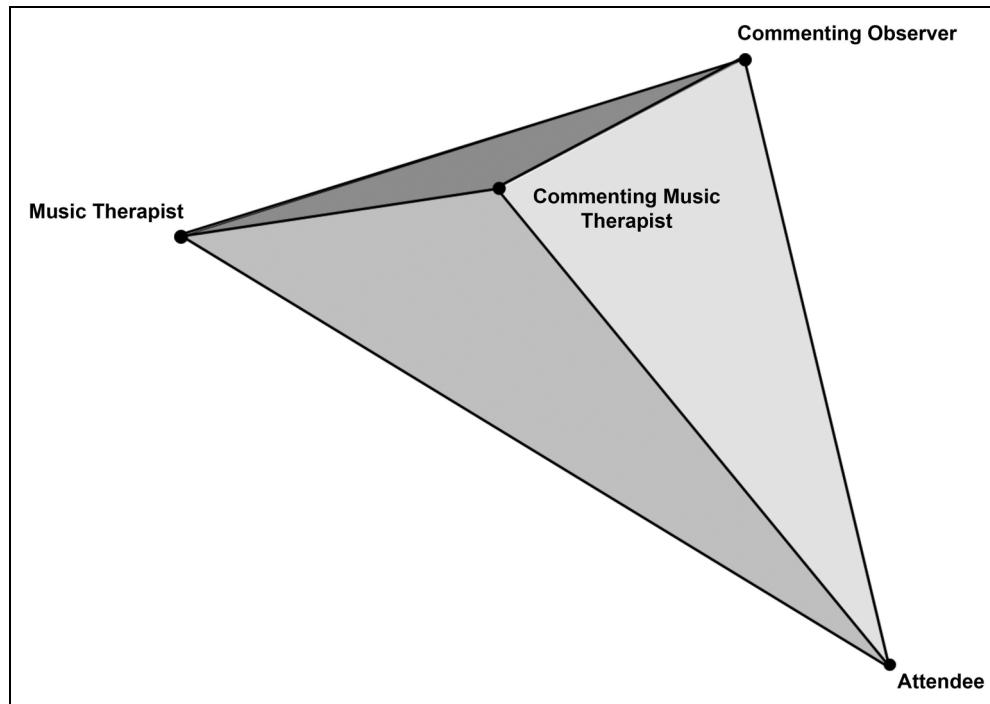


Figure 9. Tetrahedron representing averaged kappas between the four participants across the four sessions. Kappas above .16 are represented with a distance of 2, kappas between .09 and .159 are represented with a distance of 3, and kappas below .09 are represented with a distance of 4.

agreement across the four sessions (kappas in the last row in Table 9). The shape is more of a tall pyramid than an equal-sided one (though not nearly as extremely tall as Figure 8b), with the attendee as the outlier, and with the two music therapists and the two session outsiders (the commenting music therapist and the commenting observer) closest.

Figure 10 shows the tetrahedra corresponding to each session. Shapes for two of the sessions (1 and 3) are identical, although different edges are dotted, visualizing that different kappas in those sessions were statistically significant. Session 4's tetrahedron is shaped similarly to those for Sessions 1 and 3, but in this session the commenting non-therapist observer is the participant whose judgment is further away from the others rather than the attendee. Session 2's tetrahedron stands out as shaped particularly differently than the others, because in this group the attendee and commenting non-therapist observer are particularly close in judgment, and the music therapists' judgments are further apart than in the other sessions.

The visualizations make clear that in none of the sessions is there equal agreement across all parties (as a glaring salience hypothesis would predict); that even if the music therapists' judgments are closer to each other on average, this pattern is not the same in each session; and that attendees' judgments of what happened are often farther from the others' judgments.

This pattern of kappas is consistent with a moderate version of the *training leads to thinking similarly* hypothesis. It

is not at all consistent with the *session participation leads to thinking similarly* hypothesis. Rather, it is consistent with the *listeners as outsiders* hypothesis that *nonparticipation* in a session leads to thinking similarly. It is also not consistent with the evidence from the other measures that untrained outsiders think particularly differently than the trained and session participants. Rather, the attendee and the session therapists seem to think least similarly of all.

Discussion

The findings in this study are not at all consistent with a glaring salience hypothesis that everyone will see the same moments in a music therapy session as worthy of commentary or have the same interpretation of them.⁸ Across the four sessions, no moment was selected as worthy of commentary by all study participants. Also inconsistent with a glaring salience hypothesis is the fact that levels of overlap in judgment across the participants in this study were not high, in only a few cases reaching a level of "fair" agreement.

Within the fairly low levels of agreement observed here, the findings are generally consistent with the hypothesis that *music therapy training leads to thinking more similarly* (Aldridge & Aldridge, 2008; Bell et al., 2014). In this data set, the music therapists selected more of the same moments as worthy of commentary—more of the moments they selected overlapped at all, and more seconds in the

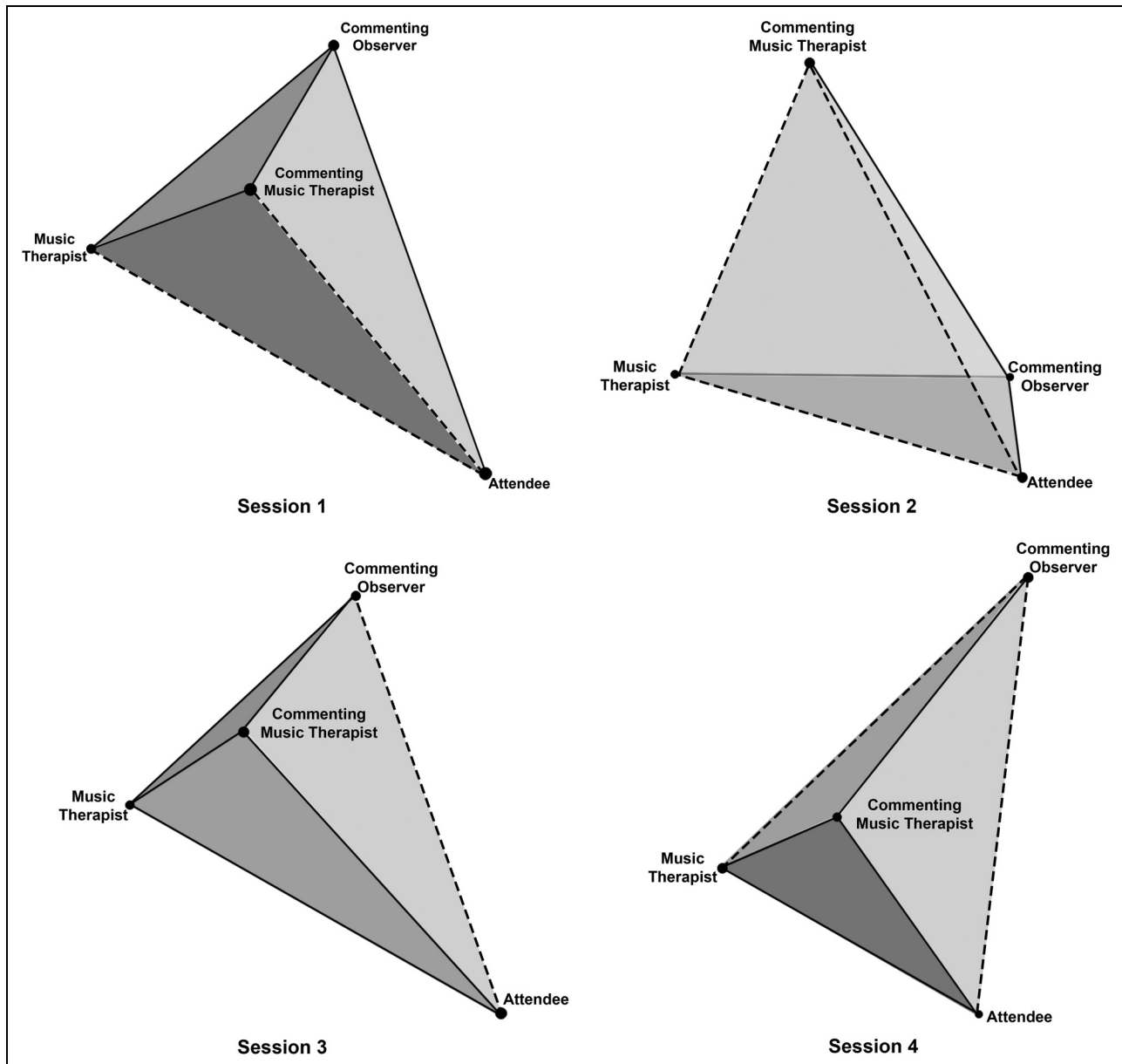


Figure 10. Tetrahedra representing kappa relationships for each of the four sessions according to the stylized scheme from Figure 8. Kappa values that are not statistically significantly different from 0 are represented with a dotted line.

moments they selected overlapped—than did the untrained observers. The session music therapists most endorsed (gave higher ratings to) statements by the commenting music therapists, and the session and commenting music therapists also averaged the highest levels of interrater agreement across all the statements they rated, in two sessions reaching the level of “fair” agreement (the most observed in this study). But the data also suggest an asymmetry in music therapists’ thinking similarly: The commenting music therapists did *not* endorse statements authored by the session music therapists more than statements by other participants.

The hypothesis that participation in a session leads to thinking more similarly is largely unsupported in this data set. By one measure, the session participants did think more similarly: The session participants selected the same moments as worthy of commentary as often as did both music therapists, and more often than the commenting observers. But the other two measures are not consistent with the hypothesis. While session attendees endorsed statements authored by the session music therapists more than statements by the commenting observers, they also endorsed statements by the commenting music therapists at the same level. And the session music

therapists did not endorse the session attendees' statements the most. Particularly contrary to the hypothesis, the levels of agreement in the session participants' ratings were lowest of all possible combinations. This is the exact opposite of what a participation hypothesis would predict.

The findings instead support a *session outsiders think more similarly* hypothesis: Outsiders who had not participated overlapped more with each other in their judgments than those who had participated in the interaction. In particular, commenting music therapists selected the same moments as the commenting outsiders more often than did the session participants. And across all the ratings the levels of agreement between commenting music therapists and the commenting observers were only a little below the levels of agreement between the two music therapists. This pattern of findings replicates evidence from prior cases studies: Listeners' judgments of characterizations of what happened in three jazz improvisations by the same professional performers overlapped more with ratings by a commenting outsider than with judgments by either participating performer (Schober & Spiro, 2016), and non-participating listeners to a classical chamber studio class performance thought more similarly about the performance than the participating performers did (Spiro & Schober, 2021). Perhaps a "view from outside" allows third-person insight that keeps music critics and marital therapists in business.

Reflections on the Method

The method implemented here—quantitative assessment of raters' levels of agreement with anonymized qualitative characterizations of what happened in a collaborative music-making setting—is resonant with other approaches to the study of music therapy that combine qualitative case study and quantitative methods to reflect and honor the individual complexity of each music therapy case while still assessing trends across multiple cases (Rickson, 2022). We see a particular advantage of this method as allowing views into discrepancies and alternative interpretations of the same moments, which may not be as visible in quantitative ratings of standardized statements designed to apply to any music therapy session. We also see as a benefit of the method that it elicits enough qualitative content to allow informative quantitative analyses even with a relatively small number of sessions and participants.

That said, the method relies on participants' verbalizations of material that may not all be straightforwardly verbalizable, which means that the method may leave out important aspects of participants' thoughts and experience as well as verbalizable comments that might remain unspoken because they seem so obvious. It also excludes participants with verbal limitations—a notable population that music therapy addresses—from being a part of such a study.

In our implementation, we gave participants the option but did not require them to watch the full session videos (to reduce burden and in the case of the session participants to elicit what had most struck them), and it is unknown how different the findings might have been if all participants had seen and heard the entire session videos.

It is important to note that our different measures give different entry points into what "thinking similarly" about a musical interaction (or an interactive event more generally) means. What you notice—which moments are salient—relative to what others notice is one aspect, but this is logically distinct from what you think about particular moments; agreeing that a moment is of interest is not the same as agreeing with what someone else thinks about that moment. As our data show—and as seen in previous studies that use multiple measures of overlapping cognition about musical interactions (Pras et al., 2017; Spiro & Schober, 2021) and about museum displays (Spiro et al., 2025), which parties think similarly by one measure may look different by a different measure.

Potential Extensions and Applications

This study selected a very particular population of participants in improvisation sessions with music therapists with particular training, and the sessions were not actually initial music therapy sessions (we did not want to interfere with therapeutic processes for participants who had self-selected or been referred to music therapy). This means that our participants' motivations may have differed from those that actual musical therapy participants bring to sessions, which can involve prior goal-setting between clients or their families and the music therapist. Our session attendees and commenting observers were all musically trained, highly verbal, and held graduate degrees, which may make their perspectives different from those of less musically trained or verbally skilled music therapy participants or observers. And of course the study involved only 16 participants and four music therapy improvisation sessions; how far these findings would generalize with a greater number of sampled sessions or more observers per session is unknown.

We see a number of different populations that should be included in this kind of study, when possible, to test the generality of the findings. In particular, we propose that it would be useful to carry out additional studies that include

- time-limited and long-term music therapy dyads
- attendees and observers with different backgrounds (musical, cultural) and demographic characteristics (age, gender, ethnicity) than these participants had
- music therapy participants referred or self-referred for the broad range of reasons that bring people to music therapy—for example, developmental, psychosocial, emotional, physical, sensory, communication, or rehabilitation issues, as well as end-of-life care

- music therapy approaches that involve different balances between music making and talking, beyond the music-centered Nordoff–Robbins approach (see McFerran et al., 2023)
- supervising music therapists and music therapy trainees

We also see possible extensions that explore the extent to which overlapping cognition correlates with other measures of music-making participants' thoughts, feelings and behaviors, for example how connected they feel with their musical partners (Spiro et al., in press).

We speculate that findings based on the method implemented here may have application in music therapy supervision and training, in that the method provides a framework for individualized reflection minimizing judgment from potentially disapproving others and focuses on the content of observations without attribution. As one music therapist in this study commented in writing, “The survey helped me gain insight in the process taking place during the session and realizing there were clearer and more frequent clinical interventions that made a difference in relating with the participant than I remember.” It may be that it is helpful for both supervisors and trainees to be further sensitized about how their perspectives may differ, and that articulating for themselves their impressions independently prior to meeting could bring additional perspectives that might not as easily emerge in time- and power-dynamic pressured supervisory conversations.

We also speculate that awareness of these findings—the fact that different stakeholders can have very different takes on what is occurring—may be helpful for music therapists in collaborative goal setting with therapy participants and family members. We do not assume that it is essential for music therapy clients or family members to understand what is happening in the same way as the therapist does—perhaps that is not even a desirable goal—but knowing more about the potential for differences in thinking may be useful. To extrapolate from findings in other arenas of interaction, there may be times where thinking differently about what is happening or has happened may lead to problematic misinterpretation that has consequences, but there may also be times where undetected conceptual misalignment can lead to productive outcomes (Schober & Dolgin, 2025).

Conclusion

This study provides new evidence on the extent to which participants in and observers of music therapy improvisation sessions agree about what happens in those sessions. The findings advance understanding about music and communication more broadly, consistent with Spiro and Schober's (2014) call for new research that brings together music therapy and music psychology frameworks, and consistent with the goal of highlighting diverse perspectives to expand

understanding of musical care (Spiro et al., in press). While none of the levels of agreement observed in this study suggest extremely similar thinking between any of the participant roles, the evidence does suggest that music therapy training leads to some shared understanding. It also suggests that an outsider perspective on an interaction involving music therapy can be more similar than an insider perspective.

Acknowledgments

We thank the participants in this study and Nordoff–Robbins Music Therapy, London where the data were collected while the first and third authors worked there. We thank Giorgos Tsiris for consultation on music therapy literature.

ORCID iDs

Neta Spiro  <https://orcid.org/0000-0002-5437-2227>

A. C. Harkness  <https://orcid.org/0009-0000-3043-6786>

Katie Rose Sanfilippo  <https://orcid.org/0000-0003-2236-3307>

Michael F. Schober  <https://orcid.org/0000-0003-1531-6326>

Informed Consent

All participants provided written consent before participating in the study.

Consent for Publication

All participants provided written consent for publication of their data.

Ethical Approval

Ethical approval for this study was granted by the Nordoff–Robbins Research Ethics Committee on October 8, 2017.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a grant from the Music Therapy Charity, UK and by faculty research funds from The New School for Social Research. Publication was supported by the Royal College of Music, London.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability

All data that can be shared are provided in this article. A full data set is not shared to retain participant anonymity.

Generative AI Tools

Generative AI tools were used to experiment with visualization options for the tetrahedra in Figures 8–10.

Redundant Publication

These data have not been published elsewhere.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Not all types of music therapy practice—of which there are many (McFerran et al., 2023; Spiro et al., 2014)—prioritize video clips.
2. See www.youtube.com/@NordoffRobbinsUK for examples of how music therapy video clips have been presented.
3. We use the term “outsiders” to refer to those not participating in a musical interaction. Note that “outsider” has also been used to describe the perspective of those who have not themselves or indirectly experienced a particular diagnosis (e.g., Rickson, 2022).
4. We instructed study participants to watch as much of the video as they chose (rather than requiring them to watch the full session) to allow session participants who already had a memory of what they thought were the interesting moments to focus on them and reduce participation burden as much as possible, while also allowing those who weren’t already sure what they thought to view the entire session if they wished. The fact that the moments participants selected came from different points in the sessions suggests that they did consider full sessions, and they did not default to commenting only on initial moments they viewed.
5. These labeling procedures were intended to make the different roles unambiguous for study participants who would be responding to the questionnaire. Note that this usage in the questionnaire is different from that in the current article, where for clarity given our research questions we use “session participants” to refer both to the attendees and the music therapists who were in the room.
6. To avoid any potential for compromising study participants’ anonymity, we are not publishing the full set of statements in the questionnaires or the comments as originally worded.
7. Following procedures in prior studies, we also calculated kappas with the more stringent requirement that rating values needed to coincide exactly, either including ratings of 6 and 7 or treating ratings of 6 and 7 as missing. As in the prior studies, using these other methods almost no kappas were significantly different from 0, and so we report the method that most clearly shows evidence of agreement.
8. Exemplifying this, the moment in the video clip at the opening of this article was selected by only one participant (the commenting music therapist). One participant (the commenting observer) strongly disagreed with 2 of the 11 statements about it and another (the session music therapist) was neutral about 2 others.

References

- Aigen, K. (2005). *Music-centered music therapy*. Barcelona Publishers.

- Aldridge, G., & Aldridge, D. (2008). *Melody in music therapy: A therapeutic narrative analysis*. Jessica Kingsley Publishers.
- 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>
- Brown, S., & Pavlicevic, M. (1996). Clinical improvisation in creative music therapy: Musical aesthetic and the interpersonal dimension. *The Arts in Psychotherapy*, 23(5), 397–405. [https://doi.org/10.1016/S0197-4556\(96\)00033-0](https://doi.org/10.1016/S0197-4556(96)00033-0)
- Bruscia, K. E. (2014). *Defining music therapy* (3rd Ed.). Barcelona Publishers.
- Clark, H. H. (1996). *Using language*. Cambridge University Press.
- Fleiss, J. L. (2003). The measurement of interrater agreement. In J. L. Fleiss, B. Levin, & M. C. Paik (Eds.), *Statistical methods for rates and proportions* (3rd Ed., pp. 598–626). John Wiley and Sons. <https://doi.org/10.1002/0471445428.ch18>
- Flower, C. (2014). Music therapy trios with child, parent and therapist: A preliminary qualitative single case study. *Psychology of Music*, 42(6), 839–845. <https://doi.org/10.1177/0305735614547065>
- Gavrielidou, M., & Odell-Miller, H. (2017). An investigation of pivotal moments in music therapy in adult mental health. *The Arts in Psychotherapy*, 52(February), 50–62. <https://doi.org/10.1016/j.aip.2016.09.006>
- Goodman, K. D. (2011). *Music therapy education and training: From theory to practice*. Charles C Thomas Publisher.
- Goodwin, C. (1986). Audience diversity, participation and interpretation. *Text*, 6(3), 283–316. <https://doi.org/10.1515/text.1.1986.6.3.283>
- Hadar, T., & Aigen, K. (2025). A theory of time in music therapy: A model for analyzing Nordoff–Robbins clinical improvisation. *Journal of Music Therapy*, 62(2), 1–31. <https://doi.org/10.1093/jmt/thaf014>
- Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116(1), 26–37. <https://doi.org/10.1037/0096-3445.116.1.26>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Ledger, A. (2015). Developing new posts in music therapy. In J. Edwards (Ed.), *The Oxford handbook of music therapy* (Vol. 1, pp. 875–893). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199639755.013.18>
- McFerran, K., Chan, V., Tague, D., Stachyra, K., & Mercadal-Brotons, M. (2023). A comprehensive review classifying contemporary global practices in music therapy. *Music Therapy Today*, 18(1), 466–485.
- Pavlicevic, M. (1995). Interpersonal processes in clinical improvisation: Towards a subjectively objective systematic definition. In T. Wigram, B. Saperston, & R. West (Eds.), *The art and science of music therapy: A handbook* (1st ed, pp. 167–178). Routledge.

- Pavlicevic, M., O'Neil, N., Powell, H., Jones, O., & Sampathianaki, E. (2014). Making music, making friends: Long-term music therapy with young adults with severe learning disabilities. *Journal of Intellectual Disabilities*, 18(1), 5–19. <https://doi.org/10.1177/1744629513511354>
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, 27(2), 169–190. discussion 190–226. <https://doi.org/10.1017/S0140525X04000056>
- Pras, A., Schober, M. F., & Spiro, N. (2017). What about their performance do free jazz improvisers agree upon? A case study. *Frontiers in Psychology*, 8, 966. <https://doi.org/10.3389/fpsyg.2017.00966>
- Rickson, D. (2022). *Music therapy with autistic children in Aotearoa, New Zealand: Haumanu ā-Puoro mā ngā Tamariki Takiwātanga i Aotearoa*. Palgrave Macmillan.
- Schegloff, E. A. (1987). Some sources of misunderstanding in talk-in-interaction. *Linguistics*, 25(1), 201–218. <https://doi.org/10.1515/ling.1987.25.1.201>
- Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123–164). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410607348>
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21(2), 211–232. [https://doi.org/10.1016/0010-0285\(89\)90008-X](https://doi.org/10.1016/0010-0285(89)90008-X)
- Schober, M. F., & Dolgin, R. S. (2025). Lost in virality: How social media can amplify hidden misinterpretations. *Technology, Mind, and Behavior*, 6(4), 298–309. <https://doi.org/10.1037/tmb0000174>
- Schober, M. F., & Spiro, N. (2014). Jazz improvisers' shared understanding: A case study. *Frontiers in Psychology*, 5, 808. <https://doi.org/10.3389/fpsyg.2014.00808>
- Schober, M. F., & Spiro, N. (2016). Listeners' and performers' shared understanding of jazz improvisations. *Frontiers in Psychology*, 7, 1629. <https://doi.org/10.3389/fpsyg.2016.01629>
- Spiro, N., Harkness, A. C., Rigamonti, A., Rossi Rognoni, G., & Schober, M. F. (2025). New methods for measuring shared and divergent visitor experience: A pilot study at the Royal College of Music Museum. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 69(1), 1016–1021. <https://doi.org/10.1177/10711813251357919>
- Spiro, N., McConnell, B. B., & Sanfilippo, K. R. M. (in press). Editorial: Interdisciplinary and intercultural perspectives on musical care. *Music & Science*.
- Spiro, N., & Sanfilippo, K. R. M. (eds.) (2022). *Collaborative Insights: Interdisciplinary perspectives on musical care throughout the life course*. Oxford University Press.
- Spiro, N., Sanfilippo, K. R. M., McConnell, B. B., Pike-Rowney, G., Bonini Baraldi, F., Brabec, B., et al. (2023). Perspectives on musical care throughout the life course: Introducing the Musical Care International Network. *Music & Science*, 6, 20592043231200553. <https://doi.org/10.1177/20592043231200553>
- Spiro, N., & Schober, M. F. (2021). Discrepancies and disagreements in classical chamber musicians' characterisations of a performance. *Music & Science*, 4, 20592043211011091. <https://doi.org/10.1177/20592043211011091>
- Spiro, N., Schober, M. F., Harkness, A. C., Waddell, G., & Perkins, R. (in press). Mothers in online songwriting groups who interpret what happened more similarly feel more connected. *Psychology of Music*.
- Spiro, N., Tsisis, G., & Pavlicevic, M. (2014). Music therapy models. In W. F. Thompson (Ed.), *Music in the social and behavioral sciences: An encyclopedia* (pp. 772–774). Sage Publications. <https://doi.org/10.4135/9781452283012.n260>
- Turley, A., Palumbo, A., Low, M. Y., Birnbaum, J. C., Turley, H., & Spellmann, M. (2026). The music engagement scale: An instrument to measure engagement in music making. *Journal of Music Therapy*, 63(1), 1–38. <https://doi.org/10.1093/jmt/thag003>