



OPEN ACCESS

EDITED BY

David Mosse,
SOAS University of London,
United Kingdom

REVIEWED BY

Tomi Bergström,
University of Jyväskylä,
Finland
Wendy J. Post,
University of Groningen,
Netherlands

*CORRESPONDENCE

Melissa Lotmore
✉ Melissa.lotmore.16@ucl.ac.uk

SPECIALTY SECTION

This article was submitted to
Psychology for Clinical Settings,
a section of the journal
Frontiers in Psychology

RECEIVED 10 September 2022

ACCEPTED 07 December 2022

PUBLISHED 06 January 2023

CITATION

Lotmore M, Ziedonis D, Alvarez
Monjaras M, Hopfenbeck M, Razzaque R,
Wilson E and Pilling S (2023) Development
and refinement of the open dialog
adherence protocol in complex mental
health care.
Front. Psychol. 13:1041375.
doi: 10.3389/fpsyg.2022.1041375

COPYRIGHT

© 2023 Lotmore, Ziedonis, Alvarez
Monjaras, Hopfenbeck, Razzaque, Wilson
and Pilling. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Development and refinement of the open dialog adherence protocol in complex mental health care

Melissa Lotmore^{1*}, Douglas Ziedonis², Mauricio Alvarez Monjaras¹, Mark Hopfenbeck³, Russell Razzaque⁴, Emily Wilson¹ and Stephen Pilling^{1,5}

¹Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom, ²Department of Health Sciences, The University of New Mexico, Albuquerque, NM, United States, ³Department of Health Sciences, Norwegian University of Science and Technology, Trondheim, Sør-Trøndelag, Norway, ⁴North East London NHS Foundation Trust, London, United Kingdom, ⁵Camden and Islington NHS Foundation Trust, London, United Kingdom

Introduction: Open dialog (OD) is a both a therapeutic practice and a service delivery model that offers an integrated response to mental health care through mobilizing resources within the service user's family and community networks through joint network meetings. Therapist adherence is a crucial to the effective delivery of interventions. A key way to measure this is through structured observation tools.

Aims: The aim of this research project is to develop and refine the Dialogic Practice Adherence Scale, for use in OD research trials in the United Kingdom.

Methods: This study was a mixed methods approach to the development of an OD practitioner adherence measure. Initial steps involved meetings and discussions with experts and a review of the literature. Content validation studies were completed using a modified Delphi technique. To assess reliability of the measure, OD network meetings were audio-recorded, and tapes were rated by two independent researchers. Inter-rater reliability and internal consistency were assessed through quantitative approaches assessing variance.

Results: Results provide a description of how the OD Adherence Manual was developed in collaboration. Validation surveys showed high levels on consensus among experts in the field on the key elements of OD network meetings. Inter-rater reliability for the total score was excellent and internal consistency analyses suggest the scale is highly reliable.

Discussion: The scale presented here is an initial attempt at rating practitioner adherence in OD network meetings. It provides encouraging evidence that this can be done with strong validity and reliability and can be completed by a range of raters with varying levels of clinical experience.

KEYWORDS

open dialogue, mental health, adherence, reliability, protocol, measure

1. Introduction

At present in England, there is excessive pressure on psychiatric inpatient beds attributed to increased demand. This takes place in the context of reduced community resources, limitations in crisis response and decreasing availability of long-term community support (Wheeler et al., 2015). Individuals suffering from complex mental disorders, defined as emotional, cognitive, or behavioral disturbances that have reached a threshold that causes substantial functional impairment are most likely to be occupying these beds (Leichsenring and Rabung, 2008; Public Health England, 2018). These disorders have a long-term impact on the individual diagnosed and their support network and often require extensive interventions and multidisciplinary or multiagency team working (Horn, 1965; Keene, 2008).

Interventions that target the social network may have a role in ameliorating mental health crises, reducing the likelihood of relapse and therefore, help to decrease pressure on inpatient psychiatric beds (Hoult et al., 1983; Olivares et al., 2013). Although Community Recovery Home Treatment Teams (CRHTTs) often acknowledge and, may attempt to work with the social network of the person in crisis, the often-limited nature of CRHTT contact and poor coordination of services militates against this. Despite the early promise shown in randomized control trials (RCTs; Johnson et al., 2005) research suggests that CRHTTs may no longer be associated with a reduction in hospital admissions (Jacobs and Barrenho, 2011). This could be due to a considerable atrophy of the key functions of CRHTT with many services offering limited home visits outside of office hours and only 50% of services providing post-hospital discharge care (Wheeler et al., 2015).

Current service responses to these problems include the development of alternatives to admission (e.g., Crisis Houses; Lloyd-Evans et al., 2014), increased capacity for psychiatric assessment in Emergency Departments, and research aimed at improving CRHTT functioning [e.g., CORE program grant led by Johnson (2013)].¹ However, these initiatives focus primarily on the management of the crisis and its aftermath, not the wider system change (e.g., continuing community support) which needs to be addressed if bed pressures are to be reduced and outcomes for service users improved in the longer term.

Epidemiological research implicates poor social networks in both the development and maintenance of mental disorder (Giacco et al., 2012). Interventions which target the social network have been advocated by developers of crisis services (e.g., Hoult in London in the 2000s) but given the brief nature of CRHTT contacts, limited staff knowledge and skills, and lack of continuity of care, such interventions are not currently provided. In addition, the evidence describing the content of these interventions, and how services which deliver them may be provided by the NHS is limited. One such model which may provide an alternative approach to crisis care is open dialog

(OD). This approach explicitly focuses on bringing about change in the social network while supporting an individual through a mental health crisis. In depth exploration of the content of this approach is required for its potential implementation into the NHS.

Developed in Finland, OD is both a therapeutic practice and a service delivery model. It offers an integrated response to mental health care with an emphasis on mobilizing resources within the service user's family and community networks through joint network meetings (Seikkula et al., 2006, 2011). Network meetings are the core therapeutic intervention within the OD approach and often take place in service users own homes. In these network meetings, service users and their networks engage in shared decision making with professionals to deploy appropriate interventions (psychological, pharmaceutical, and/or social) with the aim of developing longer term mutual support. The development of an integrated OD approach to the provision of mental health services offers the possibility of an effective alternative to the current functional model where particular functions (e.g., crisis interventions, longer-term community support) are provided by separate teams.

A systematic review by Freeman et al. (2019) found 23 studies of OD (mixed methods, qualitative, and quantitative). The review suggests that although findings of these studies have been promising the evidence is low quality and RCTs are needed to draw any additional conclusions. Uncontrolled studies report reductions in bed usage and improved recovery rates following OD interventions (Seikkula et al., 2011). Although promising, there is no high-quality evidence to support an NHS-wide adoption of this model. In order to determine whether OD is an effective alternative to the current model, the ODESSI program grant will undertake a multisite randomized control trial (RCT) comparing OD with treatment as usual (TAU). Findings from this RCT will influence whether or not changes are made more globally to NHS service structure to include more social network approaches. An important part of this research involves understanding what takes place in OD network meetings and how this links to therapeutic change.

The central component of an OD network meeting is a dialogic interaction, in which the basic feature is that each participant feels heard and responded to. Being an OD practitioner involves being able to listen and adapt to the particular context and language of every exchange and it is not possible to make specific recommendations for sessions in advance (Olson et al., 2014). However, there are distinct elements on the part of the therapists that generate the flow of dialog which in turn helps to mobilize the resources of the person at the center of the network (Olson et al., 2014). As set out in *The Key Elements of Dialogic Practice in OD* (Olson et al., 2014), there are 12 key elements or "fidelity criteria" of dialogic practice which are important for understanding the OD model (presented in Figure 1). These elements describe ways in which the practitioners can use utterances to generate new narratives amongst network members and move away from problem saturated interactions.

¹ <https://www.ucl.ac.uk/core-study>

1. Two or more therapists in the network meeting
2. Participation of clients' family and network
3. Using open-ended questions
4. Responding to clients' utterances
5. Emphasising the present moment
6. Eliciting multiple viewpoints
7. Use of a relational focus in the dialogue
8. Responding to problem discourse or behaviour in a matter-of-fact style and attentive to meanings
9. Emphasizing the clients' own words and stories, not symptoms
10. Conversation amongst professionals (reflections) in the treatment meetings
11. Being transparent
12. Tolerating uncertainty

FIGURE 1
Key elements of dialogic practice (Olson et al., 2014).

In order to ensure adequate implementation of the OD model, measures of treatment integrity such as adherence and fidelity are required. These measures will provide information to researchers and treating teams about whether or not the OD approach is being delivered as developed and intended. This is necessary to link treatment to outcome which is the wider goal of the ODDESSI RCT. The Key Elements listed above may be a useful starting point for the development of a measure of practitioner adherence within OD network meetings as they have been identified by experts in the field as integral to the OD therapeutic process.

Therapist adherence is a crucial to the effective delivery of interventions, as well as necessary to support successful dissemination across settings (Startup et al., 2002; Lange et al., 2016). It is used to reflect the degree to which therapists employ interventions prescribed by a model or framework and avoid the use of proscribed interventions during their therapeutic exchanges with service-users (Yeaton and Sechrest, 1981; Waltz et al., 1993; Schoenwald et al., 2000). The principal way that adherence is measured is through structured observation scales – measures containing the key components of a model based on its theoretical constructs. These measures must be psychometrically robust in order to accurately measure adherence and be useful for ongoing research into the efficacy of an intervention (Glasgow et al., 2005; Gearing et al., 2011). Using these measures, treatment adherence research can provide information about the successes and failures in the delivery of a model linking symptom change with therapeutic progression based on specific intervention techniques (Startup and Shapiro, 1993; Hogue et al., 1998; Onwumere et al., 2009).

Adherence scales for OD have yet to be formally developed and tested (described below). They are required for use in the ODDESSI RCT to ensure accurate implementation of the model. A measure of practitioner adherence using the key elements described above will allow researchers to more clearly establish the content of OD network meetings, ensure its successful implementation, and link the therapeutic approach with outcomes.

The “Dialogic Practice Adherence Scale” (DPAS; Olson et al., n.d.), has been developed in the United States for their healthcare

system based on expert knowledge and consensus. It is in its introductory phases and included only the 12 Key Elements and a rating scale. At present, it has not been evaluated, validated, nor has the measure been used in research trials which would subject it to rigorous reliability and validity testing. The measure requires additional development in order to determine its applicability for use in the ODDESSI research trial.

2. Aims

The aim of this research project is to develop and refine the DPAS (Olson et al., n.d.), for use in OD research trials in the United Kingdom (the ODDESSI program grant). The primary goal is to begin the process of psychometric formalization of a measure of OD practitioner adherence. This process will involve determining the essential components of the OD model, as defined by the OD Fidelity Criteria (Olson et al., 2014), developing a rating manual for the measure to allow it to be used by research staff throughout the project, and testing reliability and validity of the measure to determine its suitability for wider use.

3. Materials and methods

3.1. Design

This study is a mixed methods approach to the development of an OD practitioner adherence measure. Initial steps involved meetings and discussions with experts and a review of the literature to provide face validity. Content validation studies involved the use of surveys with results presented through narrative synthesis and summary statistics. To assess reliability of the measure, OD network meetings were audio-recorded, and tapes were rated by two independent researchers. Inter-rater reliability was assessed through quantitative approaches assessing variance.

3.2. Setting

Data for this study was drawn from the initial feasibility trial of the ODDessi work program conducted out of University College London (UCL). This is part of the initial stages of the RCT which aims to examine the implementation of OD across different NHS trusts in England and compare outcomes to TAU. The main work for this study took place at UCL with network meeting data from North East London NHS Foundation Trust (NELFT), Kent and Medway NHS and Social Care Partnership Trust (KMPT), Barnett Enfield and Haringey NHS Trust (BEH) and Devon Partnership NHS Trust (DPT). Network meetings were recorded between September 2018 and April 2019 and rating took place between January and May 2019.

3.3. Therapist and patient participants

Teams established to deliver OD interventions in the above trusts participated in this research. All practitioners (psychiatrists, psychologists, social workers, nurses and peer support workers) were trained in the OD model and integrated into practicing OD teams. Clinicians had varying degrees of training in the model, some attending training in Finland to the level of being an OD trainer themselves or more trained in the United Kingdom in a one-year foundation training or three-year full training program. Practitioners obtained written consent from all service-user trial participants and their networks for meetings to be recorded and for these recordings to be used in this research.

Service users were included in the trial if they were 18 years and above and suffering from a mental health “crisis.” Mental health “crisis” included anyone who meet criteria for referral to CRTs. There is some variability in the operational definition of “crisis” across trusts and therefore additional variability in participants presenting to services in different areas due to the makeup of the population in more rural versus urban areas. Service users were excluded from the trial if they had a primary diagnosis of dementia, primary diagnosis of a learning disability, drug and/or alcohol misuse.

A network refers to anyone closely involved in the individual service-user’s care. This includes family, friends, GPs, individual therapists, keyworkers, named nurses, members of outside agencies, etc. The service user is encouraged to identify who they would like to attend these meetings and is given the responsibility of extending these invitations on a meeting-by-meeting basis. Therefore, the make-up of each network meeting varies unpredictably in size and composition.

3.4. Raters

Five individuals were trained to use the measure and rate OD network meeting tapes. This included two highly trained OD practitioners who have a key role in the research trial and are

involved in OD training in the United Kingdom (RR and MH), a research assistant (EW) who was involved in the research trial but does not have a background in clinical or OD work. And, finally, two trainee clinical psychologists (ML and MAM) who are not trained in the OD approach but were currently undertaking DCLinPsy degrees at UCL. Raters with varying levels of background in OD were chosen in order to test whether the scale could be used by non-experts. Raters were kept blind to which practitioners were involved in the network meetings being rated, although this was not set as standard and some practitioners introduced themselves at the start of the recordings.

3.5. Survey participants

Individuals that attended the OD International Conference in London in 2018 were contacted *via* email to take part in an online survey. All individuals were actively researching or practicing OD and therefore had significant knowledge about the approach and various techniques applied in network meetings.

3.6. Procedures

3.6.1. Measure development

As a starting point, collaborators (ML, RR, and MH) met to discuss the DPAS (Olson et al., n.d.), a measure developed in the United States to measure OD adherence in network meetings. The DPAS was still in development and had not undergone any validity testing. It was used as a starting point or framework from which the research team aimed to simplify the coding process and test the protocol’s reliability and validity. The first step in the process was determining the key elements of an OD network meeting using “The Key Elements of Dialogic Practice in OD: Fidelity Criteria” (Olson et al., 2014) which set out the key methods used by practitioners in OD network meetings (presented in Figure 1). These key elements were then operationalized into specific behaviors that would be witnessable to an observer. This involved debate between the collaborators (ML, RR, MH, and SP) and four drafts were produced and open to edits.

During this process researchers in the United States (Ziedonis, Small, and Larkin) were also developing an OD adherence rating manual based on the DPAS for use in their trials. This resulted in The Dialogic Practice Fidelity Rating Manual. The Dialogic Practice Fidelity Rating Manual comprised similar components to the items that were generated through the collaboration described above. It was in draft form with a more thorough description of the elements of OD than the initial DPAS with some guidance on the process of rating and scoring an OD network meeting. However, it had not undergone any validity testing and was not being regularly or widely used. Work shifted to editing and refining this measure through consultation and debate amongst collaborators to increase the ease of use and relevance to the United Kingdom trial. Refining took place across months with

multiple drafts edited by collaborators with expert knowledge of the model. This was followed by the UCL rater training and analyses of reliability and validity.

3.6.2. Rater training

Once the coding system was agreed upon and necessary revisions made, collaborators began a series of practice trials using the measure over a two-month period. Following familiarization with the manual, all five raters individually rating 30-min to one-hour segments of one videotaped and one audiotaped OD network meeting. Following each portion rated, raters would meet and discuss scoring and increase knowledge of OD specific techniques. During this process, each individual noted specific phrases and times within the sessions that presented confusion for discussion as a group. All raters were new to using the coding system, however two were highly trained in the model and able to answer any technical questions and aid in decision making.

Following training, the five raters listened to a complete audiotaped OD session and met to discuss the completed coding criteria. Results on the criteria were visually compared for similarities and differences amongst the raters. Differences were discussed and any conflicts addressed by group consensus. Overall, agreement was established based on these initial ratings through visual inspection of the coding sheets and average ratings across the 12 items.

3.6.3. Rating

Practitioners were asked to record their network meetings with consent from the service-user and any network members present. OD sessions from different stages of treatment were included except for initial introductory sessions. There were no additional criteria that had to be met for a recording to be included in reliability analyses and, for the purposes of these analyses, it was acceptable for multiple recordings to come from the same family and same practitioners. This was because, for this study, the focus was on the utility and reliability of the measure rather than the level of adherence of the treating teams.

A 25 audio-recordings across five OD trial sites were collected for this study. Based on a literature review, this number was deemed to be acceptable and appropriate for this research (Williams et al., 2011; Pantalon et al., 2012; Gillespie, 2014; Roth, 2016). This total represented 3 audio-recordings from NELFT, 12 from KMPT, 2 from BEH, and 8 from DPT. Session length ranged from 33.02 to 115.5 min.

As this research took place in the initial pilot study stage of the RCT, no additional information was collected about service-users or practitioners other than what was on the tapes. In some sessions, introductions were made at the beginning of the recording which assisted raters in distinguishing between network member and practitioner voices. However, this was not done as standard to preserve anonymity. Therefore, it is unclear how many tapes may have been recorded by the same treating pairs or with the same network. Due to the small size of treating teams it is likely that

practitioners appeared more than once on the recordings, however, there appeared to be considerable variation in service-users and networks. As tests in this study were conducted on raters rather than therapists/families this was deemed acceptable.

Initially a random number generator was used to organize the five raters into pairs and randomly allocate the tapes for independent rating. However, as the audio-recordings were collected at different time periods from December 2018 to May 2019, audio-recordings that were collected at later dates were rated purposively by available raters.

All raters except for the primary researcher were blind to their rater pairings. Raters were not given any information about scoring until after their sessions had been submitted. The primary researcher scanned score sheets for large discrepancies (for example if one rater passed a session while another failed it) and contacted raters about these sessions. This occurred on four occasions. For training purposes, raters were requested to revisit these scores, however, at no time did they see the scoresheet of the other rater. The initial scores submitted were used in the analyses.

3.7. Analyses

3.7.1. Face/content validity

A modified Delphi technique (a method of consensus building using questionnaires) was used to gather data from respondents within their domain of expertise (Hsu and Sandford, 2007). This was done using the Qualtrics Survey Software, a free online platform for the development and data management of research surveys. Individuals with expertise in OD were contacted *via* email and sent a link to the online survey. The initial questions related to whether or not the 12 key fidelity items reflected key elements of OD practice as seen in a network meeting. Survey participants were asked to respond to this on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Respondents were then asked three further open response questions about whether they viewed these items as necessary and relevant, and whether they would make any further changes or amendments to these items. The final survey consisted of 12 Likert-response items, three qualitative feedback questions, and three respondent demographic questions.

3.7.2. Inter-rater reliability

Statistical analyses were conducted using SPSS 25. Intraclass correlation coefficients (ICCs) were calculated for all pairs of coders to estimate reliability. The convention developed by Cicchetti (1994)'s for evaluating the usefulness of ICCs was adopted for the current study and is as follows: below 0.40 = poor, 0.40 to 0.59 = fair, 0.60 to 0.74 = good, and 0.75 to 1.00 = excellent. ICC was calculated using a two-way random model with absolute agreement as per recommendations by Shrout and Fleiss (1979) for each adherence item independently as well as scale total.

3.7.3. Internal consistency

Cronbach's alpha coefficients were computed as a measure of internal consistency. A threshold of >0.70 (good) was used as a standard threshold of internal reliability (Bernstein and Nunnally, 1994). Cronbach's alpha was selected due to the use of Likert rated items in the measure. Likert items were considered on an ordinal scale in these analyses. Reliability coefficients were inspected at the item level to determine whether or not any single items significantly impacted the overall reliability of the scale.

4. Results

4.1. Measure development

The final manual was 18 pages covering the rating process and defining the key elements of OD. The retained information and descriptions enhance understanding of meaning underlying the different elements and anchor the coding framework. The anchor points describe why a rater may give a key element a certain rating. They help to distinguish a 1 (not at acceptable level), 2 (acceptable), 3 (good), and 4 (excellent). They also clearly outline when certain decisions should be made as well as the pass/fail criteria (Forsberg et al., 2015). The four-point scale was used as it had been developed in the original manual and initial comparisons showed reliability between raters with this format. Additional anchor points on the scale would have made the rating process more complex as a greater number is likely to increase the systematic variance and redundancy in a scale (Jaju and Crask, 1999).

As part of the rating process and, in line with the definition of adherence described above, it was important to get a measure of "dose" – in this case a count of specific OD-related therapeutic techniques used within the session. In order to do this, collaborators agreed it was important to rate every "utterance" made by a practitioner. This also helped to establish the proportion or monologic versus dialogic utterances and a cut-off was established regarding the necessary proportion for a session to be true to the OD model. Collaborators created a structured table with definitions of the key elements as well as monologic items. This allowed users to tally the practitioners' "utterances" to inform the subsequent ratings.

The 12 Likert-rated items on the scale reflect the 12 fidelity criteria (Olson et al., 2014; see Table 1), with each principle represented by one item. The first two items are structural and relate to the individuals in the room, i.e., number of practitioners and involvement of the network. The subsequent 10 items reflect the key therapeutic elements of the OD model. Final scores on the measure can range from 12 to 48. A score below 22 is considered to not be adherent (as this would represent more than two items rated as not at an acceptable level).

In order to rate these 12 items, the manual advises raters to refer to the tallies made within the utterance table and use these to inform their decision making. Simple presence or absence

measures were not appropriate for use in this model because OD network meetings are led by the service-user and network and, therefore, clinicians cannot be expected to engage in all OD skills at similar levels in every meeting.

At the end of the coding sheet an overall adherence rating is taken on the basis of three general questions. In order for a session to be considered adherent a score of "Yes" has to be answered on all three yes/no questions stated below.

1. Was the proportion of dialogic statements at least two-thirds (0.67)?
2. Were at least 8 of the 10 fidelity items in Section B at the level of "Acceptable" or higher?
3. Were there fewer than two instances of patronizing or disrespectful statements?

4.2. Validity

4.2.1. Face validity

A large extent of face validity of the measure was established through the parallel development process in both the United States and United Kingdom. The measure was also based on the theoretical concepts outlined by Olson et al. (2014) which provides a strong theoretical grounding based on international expert opinion.

4.2.2. Content validity

Twenty-nine individual responses were received *via* the Qualtrics Survey Software. Survey participants varied in levels of training/experience from expert >5 -years ($N=12$), advanced 2-5-years ($N=11$) and beginner <2 -years ($N=6$). All individuals were actively researching or practicing OD and therefore all had large amounts of knowledge in the area. Nine participants were primarily involved in OD research, 9 in OD practice and 11 involved in both research and practice. Participants represented an international sample (Australia = 4; Belgium = 1; Finland = 5; France = 1; Germany = 2; Italy = 1; Japan = 1; Lithuania = 2; Norway = 1; Netherlands = 2; United Kingdom = 7; United States = 1; Unknown = 1).

Results from question one of the survey are presented below in Table 1. Participants were asked "To what extent do the following items represent key elements of OD Practice as would be seen in a network meeting?" and responded on a Likert scale as described in the methods. Mean ratings for each element was above 4.0 representing agreement for all 12 items.

Participants were also asked the following open response questions: (1) What you would add to the scale? (2) What would you remove from the scale? and (3) Is there anything you would change? These questions received variable responses and are presented below (see Figures 2–4).

Overall 6 of 29 survey respondents suggested items that they would add to the scale (see Figure 2). Many of these responses

TABLE 1 Key elements survey results.

#	Key element	Mean	Min.	Max.	SD	Variance	Count
1	Two (or More) therapists in the team meeting	4.66	1.00	5.00	0.84	0.71	29
2	Participation of family and network	4.66	2.00	5.00	0.71	0.50	29
3	Ongoing use of open-ended questions throughout the treatment meeting as a way of linking client utterances and building dialog	4.38	3.00	5.00	0.67	0.44	29
4	Responding to clients' utterances: This includes responsive listening, using the clients' own words and tolerating silences in conversation	4.79	3.00	5.00	0.48	0.23	29
5	Emphasizing the present moment: Responding to immediate reactions and emotions but not interpreting or agenda setting	4.52	3.00	5.00	0.56	0.32	29
6	Eliciting multiple viewpoints: Outer and inner polyphony engaging everyone in the meeting and multiple viewpoints in an individual	4.83	4.00	5.00	0.38	0.14	29
7	Use of a relational focus in the dialog: Focus on the relational aspects of spoken stories to define relationships and elicit contextual and social information	4.24	3.00	5.00	0.68	0.46	29
8	Responding to problem discourse or behavior in a matter-of-fact style and with meaningful dialog: Seeing symptoms as "natural" responses to stressful life situations	4.41	2.00	5.00	0.77	0.59	29

(Continued)

TABLE 1 (Continued)

#	Key element	Mean	Min.	Max.	SD	Variance	Count
9	Emphasizing the clients' own words and stories, not symptoms: Help client find words to communicate more clearly, pay attention to one word or sub-sentences	4.69	3.00	5.00	0.53	0.28	29
10	Conversation amongst professionals (reflections) in the treatment meetings	4.48	3.00	5.00	0.72	0.53	29
11	Being transparent: Shared decision making. Disclosing Information on all discussions at the treatment meeting to all members present, sharing what clinicians do know and do not know	4.76	3.00	5.00	0.50	0.25	29
12	Tolerating uncertainty: No hasty judgments about symptoms, diagnosis or treatment, understanding and responding to the whole person in context rather than reacting to isolated behaviors	4.83	4.00	5.00	0.38	0.14	29

Are there any items that you would add to the scale? If so, what and why?

1. No, I think the essential moments already are in the scale.
2. Emphasizing personal ways of responding instead of "pure" professionalism.
3. Continuity, immediate response
4. Bringing yourself to the sessions, your genuine responses and owning these
5. To be open and honest about boundaries that you have or don't have in contact with somebody, so that you can be fully open to the persons and that moment, not that there is transparency about expectations of care.
6. I would add some items to assess different level of adherence between team members
7. No
8. Measures of communicative success - what is the point of being dialogical if there is no evidence that you understood them?
9. No

FIGURE 2
Items to add to the scale.

Are there any items that you would remove from the scale? If so what and why?

1. No
2. No
3. No, these are the essential moments, or let's call them "key elements" of the Open dialogue practice in the meeting.
4. No
5. No
6. N/A
7. Participation of family/social network is desirable but not necessary - many people are in crisis because of a lack of social support
8. There are too many items and many are overlapping
9. To me items 4 and 9 seem to be covering nearly the same issue - could these be combined?
10. I don't think so.

FIGURE 3
Items to remove from the scale.

Is there anything else you would change about the items on the scale?

1. No
2. No
3. No
4. I would amend the wording of item 2 as sometimes individuals do not want the network involved
5. The above fits with transparency but is more than
6. 4th and 9th items seem to express the same thing. They might be merged.
7. I would use the same Likert scale to evaluate ability and adherence. For example: the assignment of "2" in the codification of ability means "somewhat" while in the case of adherence means "fair". In our experience, this discrepancy was even more evident after data analysis. Another change I would like to propose is to use a 5 or 7 points Likert scale to make more space for critical evaluation. In fact, our impression is that the scale framed the sessions more positively than actually perceived by the raters.
8. Slightly less wordy and more helpful to define the key element
9. No
10. N/A
11. Well, the thing for me (mainly as a trainer also) is, that in different contexts it might be useful to adapt to the people in the room, or to join them from where they come. Open Dialogue rules should not be followed in a rigid way, but also flexible, dependent on the context. Whether you practice in an Institution or do home treatment, you have to be flexible. And from my Point of View there is no "One" right way to do it. Is many times a process towards. If this could be expressed also within the questions I would appreciate it. Open Dialogue is a way to more connection is not a set of rules.
12. Include importance of 1:1 sessions whether that's with a Peer, OT, nurse etc spaces are created where client can confide abuse or concerns away from the network environment where their voice maybe silenced.
13. Points 7 and 8. I think we need to be vary that the focus on "relational focus" or "problem discourse" doesn't become a "thing" or agenda...how to maintain the dialogicity and dialogical aspect throughout the whole process. For example how to honour and respect people's "problem discourse" if they find it helpful?
14. Emphasis on the conversation, and not that much on the solutions.
15. Not in this moment

FIGURE 4
Changes to the scale.

(i.e., numbers 1, 4, and 5) related to openness of response and genuineness of clinicians. Response 3 refers to an aspect of OD team structure better captured in a fidelity measure. And, response

6 advises different measures of adherence for each clinician to capture cases when one clinician may be more or less adherent than the other.

Only three of 29 respondents suggested removing any items from the scale (see Figure 3). Two of these suggested potential overlaps between items, e.g., items 4 and 9. The other response suggested decreasing the relevance of social network participation within the measure.

The final question about changes to the scale received the most responses, however, many of these responses advocated keeping the present measure (see Figure 4). One response (number 7) recommended changes in scaling used. Two (4 and 6) echoed changes advised in Figure 3 to item 2 and combining items 4 and 9. Response 12 refers to additional interventions outside of network meetings which is outside the remit of this measure. Many responses reflect the importance of clinicians being flexible and not applying specific techniques unless it fits with the nature of the current network meeting.

4.3. Scale output

Means and standard deviations for each item were computed (see Table 2). Average over all score was 33.16 out of 44 ($N=50$) showing that, overall, sites were adherent as rated on the measure. Average scores on each item ranged from adherent to good with the lowest average score on item 7 (relational focus) and the highest average score on item 4 (responsive listening).

4.4. Reliability

4.4.1. Inter-rater reliability

Inter-rater reliability for the total score was excellent. The average measure ICC was 0.906 with a 95% confidence interval from 0.785 to 0.958 [$F(24,24)=10.254, p<0.001$]. ICCs for each discrete item ranged from fair to excellent with most items in the good ($N=6$) and excellent ($N=5$) range. The one item which fell below this was item 4 (responsive listening; ICC=0.573).

4.4.2. Internal consistency

Calculation of Cronbach's alpha for the 12 items was highly reliable ($\alpha=0.848$). There was no item that could be removed from the scale to substantially increase internal consistency and all items had high item total correlations.

5. Discussion

The aim of this study was to develop and psychometrically formalize a measure of OD practitioner adherence for use in the United Kingdom-based ODESSI RCT. The initial goal of this study was to develop and refine the DPAS (in development) which had previously been developed to rate dialogic practices within network meetings. However, as the study progressed a new measure was developed, and this is presented here. Validity of the

new OD Adherence Scale has been established and internal consistency statistics report that the scale is reliable meeting the initial aims of this research project.

This is the first study to analyze the psychometric properties of the OD Adherence Scale and the results from the application of the measure provided initial adherence data which was required by NIHR in the feasibility stage of this trial. Using the scale, it was found that therapists practicing OD in the participating NHS trusts were adherent in delivery of core OD interventions. Average scores were in the adherent to good range overall and for individual items. This was true across trusts who served different populations and therefore had variability in the presentations seen within their services. It also held true with different network types and compositions.

Psychometric properties of the scale suggest that this tool may be useful in assessing adherence in OD. Modified Delphi results show that OD experts and new practitioners agree that the scale represents the key elements of the OD theoretical model. There were minimal changes suggested for the scale and many of these related to elements that would be better covered in a fidelity scale or items that are not easily operationalized for an observer rated tool. For example, individual support offered to the service user outside of network meetings would not be something observable in network meetings and would require additional interviews with service users and staff which is outside of the remit of this measure.

The use of different levels of adherence rating (adherent, good and excellent) allows the rater to make judgments about how the intervention was received by the network, whether it was appropriate, and whether or not it worked well in the context. The use of these additional rating points allows for flexibility in the sessions and addresses concerns about the rigidity of the scale described in the results. For example, neither the manual nor the measure specifies the number of occurrences of a technique for reliability. Therefore, a technique can still be rated as excellent despite occurring infrequently while another may be rated as poor in spite of occurring many times during a session. This is important for a therapeutic model such as OD with a focus on unique and flexible responses to each network in each session.

Inter-rater reliability for the overall adherence score was excellent (Shrout and Fleiss, 1979). High inter-rater reliability indicates that two randomly selected raters reliably discriminated clinician's use of and competence in different therapeutic techniques (Haddock et al., 2001) and the excellent overall score suggests that the OD Adherence Scale is a highly reliable measure. ICC ranged from fair to excellent across the items with the lowest score for item 4 responsive listening. Systematic differences between raters would likely be due to differing levels of experience both in clinical work and in OD practice. However, agreement was high for the overall score and 11 of the 12 items suggesting that training completed as part of the measure development process was sufficient, even for those with less experience with the OD model. It also shows that the measure is accessible to those with

TABLE 2 Inter-rater reliability and adherence descriptors.

Item	Description (N=25)	Mean Score	SD.	ICC
<i>Total</i>		<i>33.16</i>	<i>6.011</i>	<i>0.906</i>
<i>Avg.</i>				
1	Two (or More) therapists in the team meeting	3.06	0.682	0.612
2	Participation of family and network	2.64	0.898	0.792
3	Ongoing use of open-ended questions throughout the treatment meeting as a way of linking client utterances and building dialog	2.62	0.878	0.675
4	Responding to clients' utterances: This includes responsive listening, using the clients' own words and tolerating silences in conversation	3.16	0.889	0.573
5	Emphasizing the present moment: Responding to immediate reactions and emotions but not interpreting or agenda setting	2.68	0.891	0.824
6	Eliciting multiple viewpoints: Outer and inner polyphony engaging everyone in the meeting and multiple viewpoints in an individual	2.52	0.839	0.707
7	Use of a relational focus in the dialog: Focus on the relational aspects of spoken stories to define relationships and elicit contextual and social information	2.24	0.847	0.669
8	Responding to problem discourse or behavior in a matter-of-fact style and with meaningful dialog: Seeing symptoms as "natural" responses to stressful life situations	2.84	0.766	0.734
9	Emphasizing the clients' own words and stories, not symptoms: Help client find words to communicate more clearly, pay attention to one word or sub-sentences	3.04	0.781	0.704
10	Conversation amongst professionals (reflections) in the treatment meetings	2.58	0.835	0.727
11	Being transparent: Shared decision making. disclosing information on all discussions at the treatment meeting to all members present, sharing what clinicians do know and do not know	2.72	0.757	0.678
12	Tolerating uncertainty: No hasty judgments about symptoms, diagnosis or treatment, understanding and responding to the whole person in context rather than reacting to isolated behaviors	2.90	0.735	0.625

Italicized values are the total adherence scores.

less exposure to OD and general clinical work increasing its utility in different contexts.

The measure also demonstrated a high level of internal consistency (as reported by Cronbach's alpha) suggesting that it is a reliable measure of the intervention and that competent delivery of one individual therapeutic technique is related to competent delivery of the others (Forsberg et al., 2015). However, Cronbach's alpha is not a measure of how many constructs were measured by the scale. Additional data along with further investigation is needed to explore whether OD adherence can be efficiently rated as one global dimension.

5.1. Limitations

An important limitation of this study is the limited sample size. Significant resource is required to rate full length therapy sessions (Perepletchikova et al., 2009) and this is particularly true of OD sessions which can range from 40-min to two-hours in

length. Ideally, each of the five individual raters would have independently rated each OD tape but such resource was not available for this study. Low sample size may have contributed to variability in inter-rater reliability and internal consistency, which may have been improved with a larger sample (Shrout and Fleiss, 1979; Forsberg et al., 2015).

Additionally, there was a large time delay in receiving audio-recordings from sites which impacted the randomization process. Raters were initially randomized into pairs and to tapes but this process became purposive nearing the end of the study due to time constraints. Randomization of recordings was conducted by session, not by participant or site, therefore we had different numbers of sessions per site and there may have been some sampling bias by clinicians. As this research took place in the pilot stage of the trial, we did not collect identifying information about service users or practitioners which did not allow us to determine the impact of who was recorded on reliability outcomes. This information will be collected at later stages in the trial.

5.2. Strengths and future directions

The OD Adherence Scale is the first attempt to identify and operationalize the key elements of an OD network meeting. This study provides evidence of a consensus on the key elements of OD network meetings and dialogic practice. A strength of this research is having a varied and international team of researchers involved in the development of the measure. The parallel development processes in the United Kingdom and United States provides additional evidence of the validity of the measure. The scale presented here is an initial attempt at rating practitioner adherence in these meetings. It provides encouraging evidence that this can be done with good validity and reliability and can be completed by a range of raters with different levels of clinical experience. The scale is easy to use and does not take much longer than a network meeting to complete. It will be an important addition to OD implementation research which must report on whether OD theoretical techniques are being used adequately in practice.

This study also provides initial psychometric information as the foundation for future research and additional validation of the OD Adherence Scale. It is recommended that, as more data is collected using the measure, further analyses be performed such as those listed in the above limitations. This will improve our understanding of the measures psychometric properties providing additional evidence for or against its utility moving forward.

The manual produced as part of this research has now replaced those in development. It is being used to train raters in the United Kingdom and internationally as countries implement OD into their mental health care systems.

6. Conclusion

Perepletchikova and Kazdin (2005) propose that, in order to achieve greater scientific validity, studies looking at the relationship between fidelity and outcome should investigate empirically supported treatments, use validated fidelity measures rated by non-participant judges, and control for third variable influences. This study provides the initial element of this process for the ODDESSI program by providing psychometric information on the OD Adherence Scale.

Monitoring adherence is necessary for assessing whether participants or service users are receiving the appropriate evidence-based treatment and to identify when and how this goes wrong (Walton, 2018). It has implications for providers and wider systems and leaves us with ethical questions about how we should deliver treatment. While “perfect or near-perfect” implementation is unrealistic (Dulak and DuPree, 2008) it remains important to measure fidelity of delivery and to report on it transparently and clearly in order to translate interventions into real world settings (Walton, 2018).

Knowledge of fidelity and adherence in OD needs further development. This study is an important first step in the OD Adherence Scale’s evaluation and validation. However, the initial results presented here provide a promising foundation for the OD Adherence Scale’s utility within OD research projects.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ML wrote the text of this article and completed all surveys and statistical analyses. DZ provided a measure that he had begun development on (as described in the text). ML, RR, MA, and MH met to determine key features of OD and developed iterations of the original measure. ML, RR, MA, MH, and EW trained together to rate tapes and acted as raters and were involved in meetings to discuss and determine reliability. SP provided the idea for this research project as well as mentorship throughout and acted as an editor of this text and provided ideas for analyses. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by Open Dialogue: Development and Evaluation of a Social Network Intervention for Severe Mental Illness (ODDESSI) (Grant no. RP-PG-0615-20021).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer TB declared a past co-authorship with the author MH to the handling editor.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bernstein, I. H., and Nunnally, J. C. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6, 284–290. doi: 10.1037/1040-3590.6.4.284
- Dulak, J. A., and DuPree, E. P. (2008). Implementing matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology* 41, 327–350.
- Forsberg, S., Fitzpatrick, K. K., Darcy, A., Aspen, V., Accurso, E. C., Bryson, S. W., et al. (2015). Development and evaluation of a treatment fidelity instrument for family-based treatment of adolescent anorexia nervosa. *Int. J. Eat. Disord.* 48, 91–99. doi: 10.1002/eat.22337
- Freeman, A. M., Tribe, R. H., Stott, J. C., and Pilling, S. (2019). Open dialog: a review of the evidence. *Psychiatr. Serv.* 70, 46–59. doi: 10.1176/appi.ps.201800236
- Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., and Ngeow, E. (2011). Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. *Clin. Psychol. Rev.* 31, 79–88. doi: 10.1016/j.cpr.2010.09.007
- Giacco, D., McCabe, R., Kallert, T., Hansson, L., Fiorillo, A., and Priebe, S. (2012). Friends and symptom dimensions in patients with psychosis: a pooled analysis. *PLoS One* 7:e50119. doi: 10.1371/journal.pone.0050119
- Gillespie, M. (2014). *Revision of the multisystemic therapy (MST) adherence coding protocol: Assessing the reliability and predictive validity of adherence to the nine MST principles*. [master's thesis]. Los Angeles, CA: University of Southern California.
- Glasgow, R. E., Ory, M. G., Klesges, L. M., Cifuentes, M., Fernald, D. H., and Green, L. A. (2005). Practical and relevant self-report measures of patient health behaviors for primary care research. *Ann. Fam. Med.* 3, 73–81. doi: 10.1370/afm.261
- Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrier, N., Kinderman, P., et al. (2001). An investigation into the psychometric properties of the cognitive therapy scale for psychosis (CTS-Psy). *Behav. Cogn. Psychother.* 29, 221–233. doi: 10.1017/S1352465801002089
- Hogue, A., Liddle, H. A., Rowe, C., Turner, R. M., Dakof, G. A., and LaPann, K. (1998). Treatment adherence and differentiation in individual versus family therapy for adolescent substance abuse. *J. Couns. Psychol.* 45, 104–114. doi: 10.1037/0022-0167.45.1.104
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185. doi: 10.1007/BF02289447
- Hoult, J., Reynolds, I., Charbonneau-Powis, M., Weekes, P., and Briggs, J. (1983). Psychiatric hospital versus community treatment: the results of a randomized trial. *Aust. N. Z. J. Psychiatry* 17, 160–167. doi: 10.3109/00048678309160000
- Hsu, C. C., and Sandford, B. A. (2007). The Delphi technique: making sense of consensus. *Pract. Assess. Res. Eval.* 12, 1–8.
- Jacobs, R., and Barrenho, E. (2011). Impact of crisis resolution and home treatment teams on psychiatric admissions in England. *Br. J. Psychiatry* 199, 71–76. doi: 10.1192/bjp.bp.110.079830
- Jaju, A., and Crask, M. R. (1999). “The perfect design: optimization between reliability, validity, redundancy in scale items and response rates” in *American marketing association: Conference proceedings suppl. AMA winter educators' conference: Marketing theory*, vol. 10 (Chicago, IL: American Marketing Association), 127.
- Johnson, S. (2013). Crisis resolution and home treatment teams: an evolving model. *Adv. Psychiatr. Treat.* 19, 115–123. doi: 10.1192/apt.bp.107.004192
- Johnson, S., Nolan, F., Pilling, S., Sandor, J., Hoult, J., McKenzie, N., et al. (2005). Randomized controlled trial of acute mental health care by a crisis resolution team: the North Islington crisis study. *BMJ* 331:599. doi: 10.1136/bmj.38519.678148.8F
- Keene, J. (2008). *Clients with complex needs: Interprofessional practice*. Hoboken, NJ: John Wiley & Sons.
- Lange, A. M. C., Scholte, R. H. J., van Geffen, W., Timman, R., Busschbach, J. J. V., and van der Rijken, R. E. A. (2016). The lack of cross-national equivalence of a therapist adherence measure (TAM-R) in multisystemic therapy (MST). *J. Psychol. Asses.* 32, 312–325. doi: 10.1027/1015-5759/a000262
- Leichsenring, F., and Rabung, S. (2008). Effectiveness of long-term psychodynamic psychotherapy: a meta-analysis. *JAMA* 300, 1551–1565. doi: 10.1001/jama.300.13.1551
- Lloyd-Evans, B., Mayo-Wilson, E., Harrison, B., Istead, H., Brown, E., Pilling, S., et al. (2014). A systematic review and meta-analysis of randomized controlled trials of peer support for people with severe mental illness. *BMC Psychiatry* 14, 1–12.
- Olivares, J. M., Sermon, J., Hemels, M., and Schreiner, A. (2013). Definitions and drivers of relapse in patients with schizophrenia: a systematic literature review. *Ann. General Psychiatry* 12:32. doi: 10.1186/1744-859X-12-32
- Olson, M., Seikkula, J., and Ziedonis, D. (2014). *The key elements of dialogic practice in open dialog: Fidelity criteria*. Worcester: University of Massachusetts Medical School, 1–33. Available at: <http://umassmed.edu/psychiatry/globalinitiatives/opendialogue/>.
- Olson, M., Seikkula, J., and Ziedonis, D. (n.d.). *Dialogic practice adherence scale*. [In development].
- Onwumere, J., Kuipers, E., Gamble, C., Jolley, S., Smith, B., Rollinson, R., et al. (2009). Family interventions in psychosis: a scale to measure therapist adherence. *J. Fam. Ther.* 31, 270–283. doi: 10.1111/j.1467-6427.2009.00466.x
- Pantaloni, M. V., Martino, S., Dziura, J., Li, F. Y., Owens, P. H., Fiellin, D. A., et al. (2012). Development of a scale to measure practitioner adherence to a brief intervention in the emergency department. *J. Subst. Abuse. Treat.* 43, 382–388. doi: 10.1016/j.jsat.2012.08.011
- Perepletchikova, F., Hilt, L. M., Chereji, E., and Kazdin, A. E. (2009). Barriers to implementing treatment integrity procedures: survey of treatment outcome researchers. *J. Consult. Clin. Psychol.* 77, 212–218. doi: 10.1037/a0015232
- Perepletchikova, F., and Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice* 12:365.
- Public Health England. (2018). *Severe mental illness (SMI) and physical health inequalities: Briefing*. Available at: <https://www.gov.uk/government/publications/severe-mental-illness-smi-physical-health-inequalities/severe-mental-illness-and-physical-health-inequalities-briefing> []
- Roth, A. D. (2016). A new scale for the assessment of competences in cognitive and Behavioral therapy. *Behav. Cogn. Psychother.* 44, 620–624. doi: 10.1017/S1352465816000011
- Schoenwald, S. K., Henggeler, S. W., Brondino, M. J., and Rowland, M. D. (2000). Multisystemic therapy: monitoring treatment fidelity. *Fam. Process* 39, 83–103. doi: 10.1111/j.1545-5300.2000.39109.x
- Seikkula, J., Aaltonen, A., Alakare, B., Haarakangas, K., Keränen, J., and Lehtinen, K. (2006). Five-year experience of first-episode nonaffective psychosis in open-dialog approach: treatment principles, follow-up outcomes, and two case studies. *Psychother. Res.* 16, 214–228. doi: 10.1080/10503300500268490
- Seikkula, J., Alakare, B., and Aaltonen, J. (2011). The comprehensive open-dialog approach in Western Lapland: II. Long-term stability of acute psychosis outcomes in advanced community care. *Psychosis* 3, 192–204. doi: 10.1080/17522439.2011.595819
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420
- Startup, M., Jackson, M., and Pearce, E. (2002). Assessing therapist adherence to cognitive-behavior therapy for psychosis. *Behav. Cogn. Psychother.* 30, 329–339. doi: 10.1017/S1352465802003077
- Startup, M., and Shapiro, D. A. (1993). Dimensions of cognitive therapy for depression: a confirmatory analysis of session ratings. *Cogn. Ther. Res.* 17, 139–151. doi: 10.1007/BF01172962
- Walton, H. J. (2018). *Evaluating the implementation of interventions to improve independence in dementia*. [Doctoral dissertation]. London (United Kingdom): University College London.
- Waltz, J., Addis, M. E., Koerner, K., and Jacobson, N. S. (1993). Testing the integrity of a psychological therapy protocol. *J. Consult. Clin. Psychol.* 61, 620–630. doi: 10.1037/0022-006X.61.4.620
- Wheeler, C., Lloyd-Evans, B., Churchard, A., Fitzgerald, C., Fullarton, K., Mosse, L., et al. (2015). Implementation of the crisis resolution team model in adult mental health settings: a systematic review. *BMC Psychiatry* 15:74. doi: 10.1186/s12888-015-0441-x
- Williams, N. J., Oberst, J. L., Campbell, D. V., and Lancaster, L. (2011). The children's psychosocial rehabilitation treatment adherence measure: development and initial validation. *Community Ment. Health J.* 47, 278–285. doi: 10.1007/s10597-010-9313-5
- Yeaton, W. H., and Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. *J. Consult. Clin. Psychol.* 49, 156–167. doi: 10.1037/0022-006X.49.2.156