TILBURG UNIVERSITY

Tilburg University

Development and validation of a fidelity instrument for Cognitive Adaptation Training

Van Dam, Michelle; Van Weeghel, Jaap; Castelein, Stynke; Stiekema, Annemarie; Quee, Piotr; Kidd, Sean; Allott, Kelly; Maples, Natalie; Velligan, Dawn; Pijnenborg, Marieke; Van Der Meer, Lisette

Published in:
Schizophrenia Research: Cognition

DOI:
10.1016/j.scog.2022.100272

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Van Dam, M., Van Weeghel, J., Castelein, S., Stiekema, A., Quee, P., Kidd, S., Allott, K., Maples, N., Velligan, D., Pijnenborg, M., & Van Der Meer, L. (2023). Development and validation of a fidelity instrument for Cognitive Adaptation Training. *Schizophrenia Research: Cognition*, *31*(2), [100272]. https://doi.org/10.1016/j.scog.2022.100272

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 19. Jan. 2023

Schizophrenia Research: Cognition 31 (2023) 100272

Contents lists available at ScienceDirect

Schizophrenia Research: Cognition

journal homepage: www.elsevier.com/locate/scog

Development and validation of a fidelity instrument for Cognitive Adaptation Training

Michelle van Dam [a,b], Jaap van Weeghel [c,d], Stynke Castelein [e,f], Annemarie Stiekema [g], Piotr Quee [h], Sean Kidd [i,j], Kelly Allott [k,l], Natalie Maples [m], Dawn Velligan [m], Marieke Pijnenborg [n,o], Lisette van der Meer [a,o,*]

[a] Lentis Psychiatric Institute, Department of Rehabilitation, Zuidlaren, the Netherlands
[b] University Medical Center Groningen, Rob Giel Research Center, Groningen, the Netherlands
[c] Phrenos center of expertise on severe mental illness, Utrecht, the Netherlands
[d] Department of TRANZO, Tilburg School of Social and Behavioral Sciences, Tilburg, the Netherlands
[e] Lentis Psychiatric Institute, Research Department, Groningen, the Netherlands
[f] University of Groningen, Faculty of Behavioural and Social Sciences, Department of Clinical Psychology and Experimental Psychopathology, Groningen, the Netherlands
[g] School for Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Department of Psychiatry and Neuropsychology, Maastricht University Medical Center, Maastricht, the Netherlands
[h] GGZ Breburg, Centre for Anxiety and Mood Disorders, Breda, the Netherlands
[i] Department of Psychiatry, University of Toronto, Toronto, Canada
[j] Centre for Addiction and Mental Health, Toronto, Canada
[k] Orygen Youth Health Research Centre, Parkville, Australia
[l] Centre for Youth Mental Health, University of Melbourne, Parkville, Australia
[m] Department of Psychiatry and Behavioral Sciences, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA
[n] GGZ Drenthe, Department of Psychotic Disorders, Assen, the Netherlands
[o] University of Groningen, Faculty of Behavioural and Social Sciences, Department of Clinical and Developmental Neuropsychology, Groningen, the Netherlands

ARTICLE INFO

Keywords:
Implementation
Evidence-based practice
Fidelity
Severe mental illnesses

ABSTRACT

Purpose: Cognitive Adaptation Training (CAT) is a psychosocial intervention with demonstrated effectiveness. However, no validated fidelity instrument is available. In this study, a CAT Fidelity Scale was developed and its psychometric properties, including interrater reliability and internal consistency, were evaluated.
Methods: The fidelity scale was developed in a multidisciplinary collaboration between international research groups using the Delphi method. Four Delphi rounds were organized to reach consensus for the items included in the scale. To examine the psychometric properties of the scale, data from a large cluster randomized controlled trial evaluating the implementation of CAT in clinical practice was used. Fidelity assessors conducted 73 fidelity reviews at four mental health institutions in the Netherlands.
Results: After three Delphi rounds, consensus was reached on a 44-item CAT Fidelity Scale. After administration of the scale, 24 items were removed in round four resulting in a 20-item fidelity scale. Psychometric properties of the 20-item CAT Fidelity Scale shows a fair interrater reliability and an excellent internal consistency.
Conclusions: The CAT fidelity scale in its current form is useful for both research purposes as well as for individual health professionals to monitor their own adherence to the protocol. Future research needs to focus on improvement of items and formulating qualitative anchor point to the items to increase generalizability and psychometric properties of the scale. The described suggestions for improvement provide a good starting point for further development.

1. Introduction

Assessing treatment fidelity is considered essential in measuring the implementation of evidence-based practices (EBP) (van Weeghel, 2020). The use and rationale of fidelity assessment is twofold. First, using fidelity scales can be beneficial for clinical practice. Health professionals

* Corresponding author.
E-mail address: L.van.der.meer@rug.nl (L. van der Meer).

https://doi.org/10.1016/j.scog.2022.100272
Received 26 July 2022; Received in revised form 21 October 2022; Accepted 21 October 2022
Available online 29 October 2022
2215-0013/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

who use EBP should adhere to the critical components and procedures that have proven their effectiveness in randomized controlled trials (RCTs) to maximize the potential benefits of the EBP. To this end, fidelity instruments can be used in clinical care to monitor the degree and quality of implementation. Moreover, individual health professionals can use fidelity instruments to assess their own adherence to the protocol. In addition, program managers improving clinical services or adopting new interventions may use fidelity scales to gain a structured overview of processes that need to be set up in order to accelerate the uptake of an EBP (Rapp et al., 2008).

A second rationale to use fidelity assessments is to assess the integrity of intervention implementation for research purposes. By measuring fidelity during RCTs, researchers can evaluate whether negative outcomes are caused by poor implementation instead of an ineffective intervention (Mowbray et al., 2003). Monitoring fidelity has proven to be key in increasing successful dissemination, implementation and sustainment of EBP's in people diagnosed with SMI as demonstrated in studies for Assertive Community Treatment (ACT) and Individual Placement and Support (IPS) (Becker et al., 2014; de Winter et al., 2020; Monroe-DeVita et al., 2012). These studies showed that improving fidelity of the EBP is positively related to increased effectiveness of the intervention, which might increase the likelihood of implementation of the EBP. Thus, fidelity instruments may reflect implementation success.

Cognitive Adaptation Training (CAT) for people with severe mental illnesses (SMI) is one such intervention for which a fidelity instrument may be beneficial for both research purposes and clinical use. CAT is an EBP and focuses on reducing functional problems caused by cognitive deficits by using compensational strategies and environmental supports. Multiple trials yield evidence for the effectiveness of CAT in people with SMI (Stiekema et al., 2020; Velligan et al., 2008; Velligan et al., 2000). With CAT, a step-by-step plan is designed that is based on the needs, cognitive and behavior profile, and preferences of the service users. First, an environmental interview with the service user is performed to explore the quality of the daily life skills and to formulate the CAT goals of the service user (Velligan et al., 2010). Second, two cognitive assessments are used to determine the level of executive functioning of the service users (Nelson, 1976; Schmand et al., 2008). This is used to tailor the environmental aids to the service users. Individuals with relatively poor executive functioning benefit from compensational strategies that are more detailed and delivered at the location where the task is executed (e.g., larger and stepwise instructions in the shower that describes how to do proper personal hygiene). For individuals with moderately impaired executive functioning compensational strategies can be less detailed and placed more distant from the place where the task is executed (e.g., smaller, global instructions on the kitchen cabinet that describes a day-to-day planning). Third, an observational measure is used to assess the impact of executive functioning on behavior. A lack of executive control may result in apathy (e.g., reduced initiative and/or goal-directed behavior), disinhibition (e.g., distracted, impulsive and/or inappropriate behavior) or a combination. These behavior types provide directions for the nature of compensational strategies (Stout et al., 2003). Individuals whose behavior can be characterized as apathy may benefit from compensational strategies that prompt action (e.g., electronic devices that gives cues). For disinhibition, individuals are more supported by strategies that provide structure to diminish the distraction in the surroundings (e.g., remove off-seasonal clothing). People with a combined set of behavior types profit most from strategies that initiate action while decreasing the level of distraction once the task is being performed (Velligan, 2002).

Although the positive effects of CAT and benefits of fidelity measures for both research and clinical purposes are shown repeatedly, no validated fidelity measure is available. Therefore, the current study aimed to 1) develop a fidelity instrument based on the available literature on CAT and; 2) determine its psychometric properties.

2. Material and methods

2.1. Procedures

The CAT Fidelity Scale was developed in collaboration with research groups from the Netherlands, United States, Canada, Australia, Finland, and Belgium. Development of the scale was guided according to the procedures described by Evans and Bond (2008) and the Delphi method (Fig. 1). As a first step, the four-phased structured method by Evans and Bond was used to determine the critical components of CAT (Fig. 1, round 1). Next, the Delphi method was used to reach consensus among the items that would form the final CAT Fidelity Scale. In this method, a panel of experts provided feedback about an evolving set of items during several rounds of data collection until consensus was reached (Hasson et al., 2000). In this study four rounds of data collection were employed. Details on the development of the fidelity scale are described below.

The data used in this study is part of a large cluster RCT evaluating an innovative implementation program for CAT in long-term inpatient settings for people with SMI. Twenty-one teams from four institutions in the Netherlands participated in the study. After randomization, the teams were distributed evenly across two groups (control group and active implementation group). Both groups received training in CAT and the active implementation group received additional implementation support. The Medical Ethics Committee of the University Medical Center Groningen (The Netherlands) waived this study from ethical approval as it is not within the scope of the Medical Research Involving Human Subjects Act (file number: M17.220439). Study procedures were followed according to the Declaration of Helsinki (World Medical Association, 2013).

2.2. Development of the CAT fidelity scale

The first author (MD) constructed a list of CAT characteristics based upon a review of the scientific literature and unpublished clinical guidelines (see Appendix 1). The search strategy was performed in PubMed on April 5th 2017 with search term “Cognitive Adaptation Training”. This search revealed 20 articles which were reviewed. Items were classified at three levels: 1) CAT delivery to the service users (e.g., level of training), 2) procedural elements in delivering CAT (e.g., administered cognitive tests) and 3) organizational preconditions to provide CAT (e.g., required time).

In the first round, content experts (researchers and psychologists (n = 10)) and field experts (nurses and occupational therapists (n = 9)) were asked to participate in a four-phase survey. The content experts were recruited by email, field experts were recruited in person or via email by the content experts. All content experts are CAT trained, performed research on CAT and supervise the field experts in CAT. The field experts received training by the content experts and perform the steps necessary to set-up CAT interventions. In the first phase, all experts were asked to rate the list of CAT characteristics constructed by the first author on a 7-point Likert scale ranging from ‘not important’ (1) to ‘very important’ (7). For each item a mean score was calculated for the content experts and field experts separately. Items with a mean rating of 6 or higher by the content experts and/or field experts were selected as a concept key indicator. In the second phase, all experts were asked to select 10 items out of all CAT characteristics that they considered to be key indicators. The goal of this phase was to eliminate the tendency of respondents to label every item as “very important”. The items that were rated as “most important” by at least 50 % of the experts were identified as concept key indicators (Evans and Bond, 2008). In the third phase the experts were asked to provide an appropriate numerical value to predetermined questions, e.g., amount of time per month needed for CAT. Means were calculated for these values and were used when formalizing the CAT Fidelity Scale. In the fourth phase the experts had the option to make suggestions on key indicators that were not included in the list of items. The selected concept key indicators from the survey were then

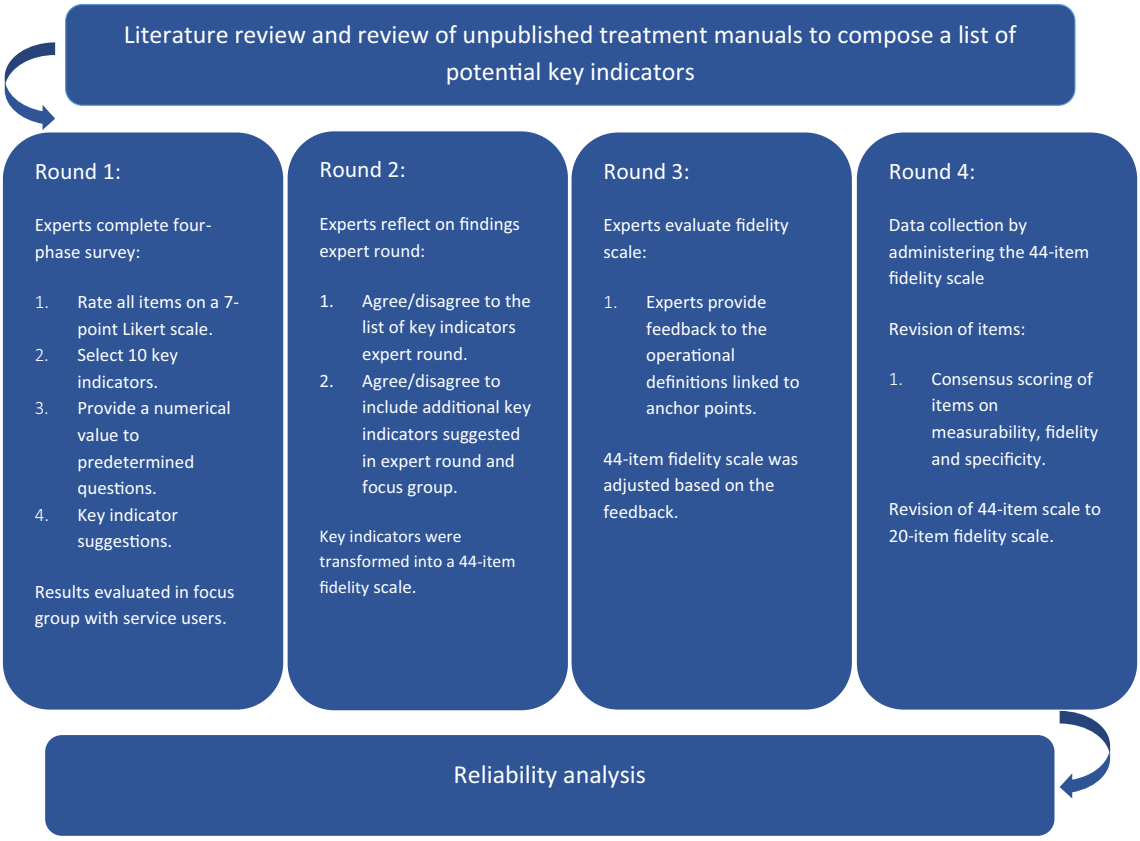

Fig. 1. Overview of the development of the CAT Fidelity Scale.

reviewed in a focus group consisting of service users who were familiar with the intervention. Based on this focus group, concept key indicators were in consensus added to the list and presented to the other stakeholders.

The second round was used to reflect on the findings of the first round and the results of the focus group meeting. All nineteen content and field experts from the first round received a list describing the concept key indicators that were excluded based upon the results of the first round and were asked if they agree (yes/no) to exclude the item from the final scale. Items that were considered as valuable by the service users in the focus group were marked. Items were excluded if >50 % agreed to exclusion. Next, the experts were presented with a second list describing the suggestions made by them in the previous round and additional concept key indicators based upon results of the focus group. They were asked if they agreed with the inclusion of the concept key indicators in the final list (yes/no) and items were included if >50 % agreed to inclusion. The primary researcher collected the data from the second round and converted the final list of key indicators in a fidelity scale, by transforming the items into operational definitions linked to anchor points on a 5-point scale (1 = no 5 = yes). The fidelity instrument was then reviewed in a focus group with the field experts delivering the intervention to ensure that the items were comprehensive and to revise the language if necessary. Items for which the wording was unclear were adjusted.

The third round was used to evaluate the fidelity scale. The experts were asked to comment on the operational definitions and anchor points of the items. Items were adjusted by these recommendations.

The fourth round was conducted after administration based upon experiences while administering the CAT Fidelity Scale (described in results section). While administering the scale, we were confronted with several issues while rating the scale. First of all, some items were difficult to score objectively, which raised the question whether they are unique the implementation of CAT or whether this holds for all EBPs. The aim of the final round was to optimize the scale by reaching consensus on each item among authors MD, LM and one other senior researcher not otherwise involved in the study. Each item was classified on three characteristics: (1) CAT-specific, (2) objectively measurable, and (3) directly related to fidelity. Items that were not CAT-specific, not objectively measurable and/or not directly related to the fidelity of CAT were excluded from the list. By adopting this participatory, consensus-driven approach, content validity of the fidelity scale was assured.

2.3. Data analysis

Inter-rater reliability was assessed through a one-day site visit by raters trained in CAT and use of the fidelity scale. Fidelity interviews (n = 73) were administered after five, eight and fourteen months and were guided by a detailed protocol, which includes instructions for preparing the visit and interview questions. Fidelity ratings were based on an interview with the persons administering CAT interventions, a review of notes describing the CAT process and, if possible, an in-home interview with a service user that received the CAT intervention. All interviews and descriptions of file notes were recorded. The interviewer and a second assessor rated the fidelity scale, the latter did so based on the audio recording. The ratings were compared and a consensus score was determined for items on which the ratings disagreed.

The data was analysed in SPSS version 26.0 (Corp, 2019). The percentage exact agreement and the kappa coefficient was calculated to determine the interrater reliability. In addition, the percentage exact agreement allowing one point difference is described to provide information on the distribution of the ratings (Kottner et al., 2011). Cohen's Kappa coefficient ranges from 0 to 1, where values of 0 indicates no agreement, 0.01–0.20 none to slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 almost perfect agreement (Landis & Koch, 1977). Cronbach's alpha was calculated to measure the internal consistency and was based on the consensus ratings. Cronbach's

Alpha values below 0.50 indicates unacceptable internal consistency, 0.50–0.59 poor, 0.60–0.69 questionable, 0.70–0.79 acceptable, 0.80–0.90 good and above 0.90 excellent (George and Mallery, 2003; Kottner et al., 2011).

3. Results

3.1. Development of CAT fidelity scale

In total, 51 CAT characteristics were subtracted from the literature review. Thirty-nine items remained after round one (18 items) and round two (21 items). An overview is presented in Table S1. These items were then transformed into operational definitions and anchor points (Supplementary S2). The items were adjusted based on the feedback provided in the round three. For example, the answer option 'not applicable' was added to the fidelity scale, as some of the items may be impossible to answer as it assumes a certain team/organizational structure that may not exist (e.g., 'The supervisor communicates with team managers at least half-yearly to ensure that CAT is integrated, to solve pragmatic issues and to actively promote the advantages of CAT'). In addition, recommendations were made regarding the wording of items, such as items describing the involvement of caregivers and family. These were transformed by changing 'caregivers and family members' to 'significant others', so that anyone who is important to the individual can included if appropriate. This resulted in the 44-item CAT Fidelity Scale (Supplementary S3). In the final round of the Delphi method, 24 items were excluded from the list based upon consensus among members of the research group (authors MD, LM and one other senior researcher). Disagreement among the items was mostly related to whether the items contribute to CAT fidelity (items: 6, 7, 15, 21, 30, 31, 37, 38, 39) and whether they are objectively measurable (items: 21, 22, 29, 30, 38, 43, 44). For example, item one: 'The CAT-specialist has an energetic and enthusiastic attitude and uses this to encourage and stimulate the service users and colleagues.' Reason to exclude this item was that although it is important to have an enthusiastic spirit to motivate service users, this is not limited to CAT but applies to recovery-oriented care in general. Furthermore, it is difficult to rate this item objectively. This also applied to item three: 'The CAT-specialist has knowledge of behaviour techniques (e.g., positive reinforcement, shaping and antecedent control) and applies these during CAT visits to motivate and support service users.' The use of behavior techniques is not limited to CAT, but is important for all recovery-oriented practices. Second, it is not one of the core elements of CAT since the absence of knowledge on behavioural techniques does not per se impact CAT fidelity. Another example is item eleven: 'Significant others have received information about CAT (e.g., basic principles, working mechanism) and have been explained how the service user may benefit from the intervention'. That is, if significant others would not receive information, this would not directly affect the treatment fidelity. The classification of the items of round four is presented in Table S4. Overall, after four rounds of consensus ratings, 20 items remained and were included in the final CAT Fidelity Scale (Supplementary S5).

3.2. Psychometric properties

3.2.1. Interrater agreement and reliability

The average interrater agreement for the 20-item CAT fidelity scale, was 69.4 % (range 52.0 % - 97.3 %) and the average interrater agreement allowing one point difference was 82.2 % (range 70.6 % - 98.7 %). The interrater reliability (Cohen's Kappa) was slight for two items (range 0.17–0.19), fair for nine items (range 0.21–0.38) and moderate for seven items (range 0.44–0.55). For two items (level of training and continuation CAT visits: outreach setting) no Kappa coefficient could be calculated because there was no variability in the ratings. All health practitioners received the same training and therefore all scores were similar. For item 'continuation CAT visits: outreach setting', all scores were rated as 'not applicable'. An overview is presented in Table 1.

3.2.2. Internal consistency

Internal consistency (Cronbach's alpha) over all 20 items was 0.91, which is considered excellent (George and Mallery, 2003).

4. Discussion

The aim of this study was to develop a fidelity scale for CAT and to assess the validity and reliability of the scale. Content validity, interrater reliability and internal consistency of the scale were examined. Content validity was assured by applying a consensus-driven approach among content and field experts of CAT in several feedback rounds. In addition, feedback was retrieved on CAT-specificity, measurability and fidelity relevance of the items after administration of the scale. Subsequently, 20 items remained on which the interrater agreement and internal consistency were determined. This 20-item scale consists of four items related to the CAT-specialist (e.g., creativity environmental aids, level of training) and 16 items related to the CAT procedures (e.g., assessment neurocognitive tests, reviewing CAT goals). The assessors had an acceptable interrater agreement with a mean of 69.4 % on the individual items of the 20-item CAT Fidelity Scale, and an interrater reliability ranging from slight to moderate. The interrater agreement improved to 82.2 % when allowing one point difference, indicating that several items may need revisions to further improve the scale to arrive at comparable interrater agreement values of other fidelity scales (88 % - 94 %; Egeland et al., 2020; Joa et al., 2020). The 20-item CAT Fidelity Scale has an excellent internal consistency, indicating that the items all measure a single construct. Altogether, the psychometric properties of the 20-item CAT Fidelity Scale are fair.

Initially 39 items were marked as core CAT principles resulting in a 44-item concept CAT Fidelity Scale, yet after administration of the scale 24 items were removed resulting in a 20-item fidelity scale. Most of the removed items were related to the organizational context in which CAT is implemented (e.g., whether there is a supervisor available to interpret the assessment findings) and were in consensus considered not to be directly related to the fidelity of CAT. For example, if a supervisor is not available to provide monthly supervision, it does not have a direct impact on the fidelity. Some of these items were also marked as not applicable by researchers from countries other than the Netherlands as it assumes a certain organizational structure that may not exist. This level of detail and inclusion of contextual factors (such as predetermined amount of time or on-spot supervision) is issue to consider when developing a fidelity instrument. Lilleleht (2005) demonstrated that highly detailed manuals and protocols may undermine the sense of autonomy of clinical workers. This concern is raised in other research as well, describing that standardization of care by clinical guidelines and evidence-based interventions may undermine the clinical experience of the health professionals (Forsner et al., 2010). Also, the adjustment of many contextual factors to adopt the EBP may cause reluctance in organizations to adopt the EBP as more investments are required, which may be a reason not to adopt the intervention. As a result, implementation may be unsuccessful. Furthermore, Bond and Drake (2020) demonstrated that scales with <25 items are optimal since they are less time consuming and therefore more likely to be adopted, which suggests that a 20-item scale accommodates the needs and possibilities of care workers better than the 44-item scale.

The current CAT Fidelity Scale has sufficient psychometric properties to be used in clinical practice and research. However, refinement of the items will likely improve the scale. For example, the interrater reliability on item eight ('The CAT-specialist collects all information of the EFA, FrSBe and cognitive tests and summarizes it in a CAT Treatment Plan') is low. A clearer description of the scale points might lead to better agreement among the raters. In the current scale, some raters scored 'no' if one out of three were present (e.g., only EFA administered), while others scored mostly no, indicating that the answering options and/or

Table 1
Mean score and inter-rater reliability per item.

Item	Mean score (SD)	Exact agreement	Agreement allowing 1 point difference	Cohen's κ	SE of κ	Strength of agreement
CAT employment						
1. Creativity environmental aids	1.22 (1.99)	77.3 %	86.6 %	0.55	0.08	Moderate
2. Flexibility environmental aids	1.03 (1.81)	68.0 %	77.3 %	0.33	0.07	Fair
3. Flexibility solutions	0.62 (1.42)	76.0 %	85.4 %	0.36	0.09	Fair
4. Level of training[a]	3.97 (0.33)	97.3 %	–	0.00	–	–
CAT procedures						
5. Assessment EFA	4.20 (1.45)	70.7 %	85.4 %	0.44	0.09	Moderate
6. Assessment FrSBe	3.36 (1.88)	65.3 %	83.9 %	0.54	0.07	Moderate
7. Assessment neurocognitive tests	3.07 (1.92)	61.3 %	82.6 %	0.51	0.07	Moderate
8. Use CAT Treatment Plan	1.36 (1.59)	37.3 %	73.3 %	0.16	0.07	Slight
9. Deciding on CAT goals	2.68 (2.30)	65.3 %	84.0 %	0.49	0.07	Moderate
10. Prioritization CAT goals	1.69 (2.19)	57.3 %	70.6 %	0.35	0.07	Fair
11. Higher order CAT goals	1.39 (1.90)	57.3 %	77.3 %	0.38	0.07	Fair
12. Reviewing CAT goals	1.15 (1.95)	64.0 %	78.6 %	0.32	0.08	Fair
13. Involvement service user	1.60 (2.01)	70.7 %	80.0 %	0.53	0.07	Moderate
14. Appropriateness environmental aids	1.15 (2.00)	70.7 %	74.7 %	0.34	0.09	Fair
15. Rationale environmental aids	0.58 (1.54)	80.0 %	82.7 %	0.19	0.11	Slight
16. Active CAT visits	3.20 (1.75)	56.0 %	78.7 %	0.44	0.07	Moderate
17. Frequency CAT visits: outreach setting	0.03 (0.20)	96.0 %	98.7 %	0.32	0.16	Fair
18. Frequency CAT visits: clinical setting	0.64 (0.89)	52.0 %	73.4 %	0.21	0.09	Fair
19. Continuation CAT visits: outreach setting[a]	0.01 (0.08)	97.3 %	98.7 %	0.00	–	–
20. Continuation CAT visits: clinical setting	0.41 (0.75)	68.0 %	90.7 %	0.28	0.09	Fair

CAT = Cognitive Adaptation Training; EFA = Environmental and Functional Assessment; FrSBe = Frontal Systems Behavior Scale; SD=Standard Deviation; SE = Standard Error.

[a] No Kappa coefficient can be calculated due to the absence of variability in the ratings.

the item wording is unclear. This also holds for double-barrelled items such as item 26 ('The external environmental aids are appropriate for both the service user's behavior type and level of executive functioning'). In addition, the most commonly noted weakness by the fidelity assessors concerned the range of the scale points – the scale points may not be distinctive enough to provide a valid and reliable score. Using the 5-point Likert scale may have caused noise in the ratings, which may have deflated the interrater reliability scores of the scale. In line with other validated fidelity scales (Bond et al., 2012; Joa et al., 2020), an approach to overcome these issues is to change the scoring options to behaviourally-anchored criteria on five discrete scale points, rather than a 5-point Likert scale ranging from no (1) to yes (5). The interrater reliability for those scales is high. The scoring options of item four ('The CAT-specialist has received CAT training by a qualified trainer. The training entailed the following elements') are anchored to specific criteria, which could serve as an example when revising the other items.

Future research needs to focus on determining cut-off points to translate the fidelity score into useful clinical interpretations. For instance, what score represents exemplary or good CAT fidelity. These issues might be difficult to formalize as the extent to which CAT can be provided is also dependent on service user characteristics. For example, some service users have difficulties in formalizing CAT goals and may come up with one single goal. The item 'Prioritization of goals' is then not applicable and causes a lower fidelity score. However, this does mean that CAT was not provided according to the intervention's model standards. More research is needed to resolve these issues, including predictive validity. Fidelity scales of other EBP have shown that higher fidelity results in better outcomes (Bond and Drake, 2020; de Winter et al., 2020; Durlak and DuPre, 2008), yet for CAT this is still unclear. This study did not assess predictive validity as this was a first attempt to develop a fidelity scale for CAT and to gain experience in the use of the scale. Future studies that revise the scale should include a measure to assess the predictive fidelity. This could be measured by applying the CAT Fidelity Scale in sites with low fidelity (e.g., new sites) and experienced sites with strong functional outcome. In this research, different sources of information were used in the fidelity ratings, namely visual presentation of notes, in-person interviews and home visits, compared to audio recording, which might have lowered the interrater reliability. Future research should consider using a uniform approach when rating the fidelity scales to limit the influence of method variance on the interrater reliability.

A strength of this study is that the scale is developed in collaboration with multiple international CAT research groups from United States, Canada, Australia, and Finland. Moreover, we included researchers, practitioners and service users in the developmental phase, which makes it a bottom-up participatory approach in which key stakeholder perspectives are represented. This study also adds to the limited research on the psychometric properties of fidelity scales. Use of the scale in clinical practice may serve a dual purpose. First of all, the scale can be used as a self-monitoring instrument to examine whether a practitioner is conforming to the standards of CAT. Second, it may be used in research, or other forms of formal monitoring, with the aim of strengthening clinical use and to inform researchers or other stakeholders in the need for extra training or support.

A limitation to this study is that the administered fidelity scale only included sites that provide care and support for people in inpatient settings in The Netherlands, while the scale has been developed in collaboration with international researchers and clinicians who provide care in a variety of care facilities. It is possible that different facets come to light when administrating the fidelity scale in other contexts, though the protocol of CAT is similar across settings. The CAT Fidelity Scale should be administered in other facilities to investigate whether the scale is generalizable and to make revisions when necessary. Particularly since other studies demonstrated that the dissemination of EBP to new settings and populations has led to modifications of the fidelity scales (Bond et al., 2019).

The current scale shows a fair interrater reliability and an excellent internal consistency which in its current form is useful for both research purposes as well as for individual health professionals to monitor their own adherence to the protocol. However, fidelity instruments are not fixed, but are subject to new insights and the continued development of the EBP. Therefore, this fidelity instrument and the described suggestions for improvement provide a good starting point for further development.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scog.2022.100272.

Funding

The study was funded in April 2018 by a grant of the Foundation for Support (*Stichting tot Steun* VCVGZ; grant number: 247).

CRediT authorship contribution statement

Each of the authors has substantially contributed to this research and drafting the manuscript. The authors have no conflict of interest.

Declaration of competing interest

None.

Acknowledgements

The authors thank all clinical care workers and service users who contributed to the development of the scale by providing data and valuable feedback. We would also like to thank the nursing staff of the mental health institutions that participated in the study for their time and effort in contributing to the fidelity assessments.

References

Becker, D.R., Drake, R.E., Bond, G.R., 2014. The IPS supported employment learning collaborative. Psychiatr. Rehabil. J. 37 (2), 79–85. https://doi.org/10.1037/prj0000044.

Bond, G.R., Drake, R.E., 2020. Assessing the fidelity of evidence-based practices: history and current status of a standardized measurement methodology. Admin. Pol. Ment. Health 47 (6), 874–884. https://doi.org/10.1007/s10488-019-00991-6.

Bond, G.R., Peterson, A.E., Becker, D.R., Drake, R.E., 2012. Validation of the revised individual placement and support fidelity scale (IPS-25). psychiatric services 63, 758–763. https://doi.org/10.1176/appi.ps.201100476.

Bond, G.R., Drake, R.E., Pogue, J.A., 2019. Expanding individual placement and support to populations with conditions and disorders other than serious mental illness. Psychiatric Services 70, 488–498. https://doi.org/10.1176/appi.ps.201800464.

Corp, B.M., 2019. IBM SPSS Statistics for Windows [computer software]. Armonk, NY.

De Winter, L., Couwenbergh, C., van Weeghel, J., Bergmans, C., Bond, G.R., 2020. Fidelity and IPS: does quality of implementation predict vocational outcomes over time for organizations treating persons with severe mental illness in the Netherlands? Soc. Psychiatry Psychiatr. Epidemiol. 55 (12), 1607–1617. https://doi.org/10.1007/s00127-020-01890-0.

Durlak, J.A., DuPre, E.P., 2008. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am. J. Community Psychol. 41 (3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0.

Egeland, K.M., Heiervang, K.S., Landers, M., Ruud, T., Drake, R.E., Bond, G.R., 2020. Psychometric properties of a fidelity scale for illness management and recovery. Administration and Policy in Mental Health 47 (6), 885–893. https://doi.org/10.1007/s10488-019-00992-5.

Evans, L.J., Bond, G.R., 2008. Expert ratings on the critical ingredients of supported employment for people with severe mental illness. Psychiatric Rehabilitation Journal 31 (4), 318–331. https://doi.org/10.2975/31.4.2008.318.331.

Forsner, T., Hansson, J., Brommels, M., Wistedt, A. A., & Forsell, Y. (2010). Implementing clinical guidelines in psychiatry: A qualitative study of perceived facilitators and barriers. BMC Psychiatry, 10, 8-8. doi:10.1186/1471-244X-10-8.

George, D., Mallery, P., 2003. SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 Update, 4th ed. Allyn & Bacon, Boston, MA.

Hasson, F., Keeney, S., McKenna, H., 2000. Research guidelines for the Delphi survey techniqu. Journal of Advanced Nursing 32 (4), 1008–1015.

Joa, I., Johannessen, J.O., Heiervang, K.S., Sviland, A.A., Nordin, H.A., Landers, M., Bond, G.R., 2020. The family psychoeducation fidelity scale: psychometric properties. Admin. Pol. Ment. Health 47 (6), 894–900. https://doi.org/10.1007/s10488-020-01040-3.

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., Streiner, D.L., 2011. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J. Clin. Epidemiol. 64 (1), 96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Lilleleht, E., 2005. Paradox in practice? The rhetoric of psychiatric rehabilitation. Int. J. Psychosoc. Rehabil. 10, 89–103.

Monroe-DeVita, M., Morse, G., Bond, G.R., 2012. Program fidelity and beyond: multiple strategies and criteria for ensuring quality of assertive community treatment. Psychiatric Services 63, 743–750. https://doi.org/10.1176/appi.ps.201100015.

Mowbray, C.T., Holter, M.C., Teague, G.B., Bybee, D.I., 2003. Fidelity criteria: development, measurement, and validation. Am. J. Eval. 24, 315–340.

Nelson, H.E., 1976. A modified card sorting test sensitive to frontal lobe defects. Cortex; a Journal Devoted to the Study of the Nervous System and Behavior 12 (4), 313–324.

Rapp, C.A., Etzel-Wise, D., Marty, D., Coffman, M., Carlson, L., Asher, D., Whitley, R., 2008. Evidence-based practice implementation strategies: results of a qualitative study. Community Ment. Health J. 44 (3), 213–216. https://doi.org/10.1007/s10597-007-9109-4.

Schmand, B., Groenink, S.C., van den Dungen, M., 2008. Letter fluency: psychometric properties and dutch normative data. Tijdschr. Gerontol. Geriatr. 39 (2), 64–76.

Stiekema, A.P.M., van Dam, M.T., Bruggeman, R., Redmeijer, J.E., Swart, M., Dethmers, M., van der Meer, L., 2020. Facilitating recovery of daily functioning in people with a severe mental illness who need longer-term intensive psychiatric services: results from a cluster randomized controlled trial on cognitive adaptation training delivered by nurses. Schizophr. Bull. 46 (5), 1259–1268. https://doi.org/10.1093/schbul/sbz135.

Stout, J.C., Ready, R.E., Grace, J., Malloy, P.F., Paulsen, J.S., 2003. Factor analysis of the frontal systems behavior scale (FrSBe). Assessment 10 (1), 79–85. https://doi.org/10.1177/1073191102250339.

van Weeghel, J., 2020. Fidelity scale: from black box to holy grail. Administration and Policy in Mental Health 47 (6), 927–931. https://doi.org/10.1007/s10488-020-01057-8.

Velligan, D. I. (2002). Cognitive adaptation training: The use of compensatory strategies in the psychosocial rehabilitation of patients with schizophrenia. (manual for therapists).

Velligan, D.I., Bow-Thomas, C., Huntzinger, C., Ritch, J., Ledbetter, N., Prihoda, T.J., Miller, A.L., 2000. Randomized controlled trial of the use of compensatory strategies to enhance adaptive functioning in outpatients with schizophrenia. Am. J. Psychiatry 157 (8), 1317–1323.

Velligan, D.I., Diamond, P.M., Mintz, J., Maples, N., Li, X., Zeber, J., Miller, A.L., 2008. The use of individually tailored environmental supports to improve medication adherence and outcomes in schizophrenia. Schizophr. Bull. 34 (3), 483–493.

Velligan, D., Maples, N., Ritch, J., 2010. Cognitive adaptation training: the use of compensatory strategies in the psychosocial rehabilitation of patients with schizophrenia (manual). UTHSCSA, Department of Psychiatry, San Antonio.

World Medical Association, 2013. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 310 (20), 2191–2194. https://doi.org/10.1001/jama.2013.281053.