

NEUROCOUNSELING

Treatment Fidelity in Neuroscience-Informed Cognitive-Behavior Therapy: A Feasibility Study

Thomas A. Field
Raissa Miller
Eric T. Beeson
Laura K. Jones

Neuroscience-informed cognitive-behavior therapy (nCBT) is an emerging approach that is being refined in preparation for efficacy trials. This feasibility study defined the essential components of the nCBT model and evaluated whether expert raters could determine if trained clinicians adhered to or deviated from the model. The study sample featured 11 licensed mental health professionals who participated in a simulated client session 8 weeks after the conclusion of a 3-day training. Sessions were recorded and reviewed by the research team, who evaluated trainee videos from 11 simulated client sessions. Interrater consistency among four raters ranged from $\kappa = .64$ to $\kappa = .84$. Considerations for model development, modifications to the fidelity scale, and recommendations for counseling researchers and practitioners regarding treatment fidelity are discussed.

Neuroscience is being integrated into counseling and psychotherapy to justify existing therapeutic models and inform the creation of new theories (Blom et al., 2014; Blom et al., 2017; Gonçalves & Perrone-McGovern, 2016; Gros, 2015). Given decades of empirical support, cognitive-behavior therapy (CBT) is referred to as the “gold standard of psychotherapy” (David, Cristea, & Hofmann, 2018, p. 1) and has demonstrated significant impacts on a variety of outcomes across disorders, including measures of brain functioning (e.g., Barsaglini, Sartori, Benetti, Pettersson-Yeo, & Mechelli, 2014). However, contemporary understandings of neuroscience counter some of the basic premises of CBT and have informed revisions to traditional CBT models (Beck & Haigh, 2014; Clark & Beck, 2010). Neuroscience-informed cognitive behavior therapy (nCBT) is one such emerging model that integrates findings from neuroscience with traditional CBT approaches (Field, Beeson, & Jones, 2015).

Thomas A. Field, Division of Graduate Medical Sciences, Boston University School of Medicine; Raissa Miller, Department of Counselor Education, Boise State University; Eric T. Beeson, The Family Institute at Northwestern University and Counseling@Northwestern; Laura K. Jones, Department of Health and Wellness, University of North Carolina at Asheville.

This study was supported by an internal seed grant from Boise State University's College of Education. Correspondence concerning this article should be addressed to Thomas A. Field, Division of Graduate Medical Sciences, Boston University School of Medicine, 72 East Concord St., Suite B-212, Boston, MA 02118. E-mail: tfield@bu.edu

Developed by counseling professionals, nCBT offers one of the first models of translational neuroscience specifically in the clinical mental health counseling field (Field et al., 2015). The model uses a semi-structured multi-phasic approach (Phase 1, assess and attend to the physiological; Phase 2, build the brain from the bottom up; Phase 3, connect the bottom to the top). This approach is grounded in a novel conceptualization framework known as the Waves of the New ABCs (Waves; Beeson, Field, Jones, & Miller, 2017). The Waves model is informed by the principles of dual information processing (i.e., top-down, bottom-up), which has been identified as an important component for emotional regulation and intervention selection (e.g., Ochsner et al., 2009; McRae, Misra, Prasad, Periera, & Gross, 2012). The Waves model provides counselors with an assessment framework to guide treatment selection based on clients' prominent ways of responding. For instance, if a client's presenting concerns seem to be generated from the bottom up (i.e., the antecedent is a sensory-based experience outside of conscious awareness), then treatment focuses on bottom-up interventions (e.g., biofeedback, somatic experiencing), whereas if the presenting concerns seem to be generated from the top down (i.e., the client is cognitively aware of the antecedent), then treatment focuses on top-down interventions (e.g., cognitive reappraisal).

According to the National Institutes of Health (NIH) model of intervention development, nCBT is currently in Stage I of development (Onken, Carroll, Shoham, Cuthbert, & Riddle, 2014). Field, Beeson, and Jones (2016) conducted a preliminary study examining counselors' and clients' perceptions of nCBT and found that counselors and clients had high levels of belief in the credibility and effectiveness of the model. Field, Beeson, Jones, and Miller (2017) conducted a follow-up study and found that counselors' understanding and allegiance to the model and their ability to communicate positive expectancy to clients were critical factors in facilitating clients' trust and engagement in nCBT. In the Field et al. (2017) study, clients' experiences of positive change using nCBT principles supported continued engagement with nCBT treatment.

Although the initial findings regarding nCBT are encouraging, they are not without limitations. The Field et al. pilot studies from 2016 and 2017 had small sample sizes and relied on measures of perceived credibility and effectiveness rather than outcome measures. Additionally, no control or comparison groups have been included in the evaluation of treatment effects. Although these limitations are notable, they are also expected given the current stage of model development, as pilot testing is essential for refining the model. The NIH's second step in intervention development is the evaluation of treatment fidelity (Onken et al., 2014).

RESEARCH INTO TREATMENT FIDELITY

Mowbray, Holter, Teague, and Bybee (2003) defined fidelity as "the extent to which delivery of an intervention adheres to the protocol or program

model originally developed” (p. 315). Melnyk and Morrison-Beedy (2012) proposed the term *process fidelity* in reference to the following of manualized procedures and the term *content fidelity* in reference to counselors’ adherence to the specific content of an intervention as described in the treatment manual. Treatment fidelity is essential in rigorous program evaluation and outcome research, lending credibility to internal validity for research and helping prevent therapist drift in clinical practice (Melnyk & Morrison-Beedy, 2012; Waller & Turner, 2016). For researchers and clinicians to claim their approach is effective, they must provide evidence that practitioners implemented the treatment consistently and in alignment with the treatment manual or protocol. Additionally, Reiser and Milne (2014) emphasized that treatment fidelity is a central component of safe and effective practice and proposed that the effectiveness of clinical supervision should be assessed on the degree to which supervisees are following treatment protocols with fidelity.

Despite the importance of treatment fidelity, we were only able to find one recent counseling-specific example in the counseling literature detailing the development and implementation of valid fidelity measures. Ray, Purswell, Haas, and Aldrete (2017) examined interrater reliability (IRR) for the Play Therapy Skills Checklist (PTSC), a form used to assess counselors’ adherence to the Child-Centered Play Therapy protocol. Ray et al. (2017) described their process of refining PTSC item definitions, a process that included convening an expert panel, reviewing historical play therapy texts and video recordings, engaging in practice sessions and discussions, and analyzing scores to assess for reliability. The original development of the checklist, however, was not described in detail.

Because of the limited guidance found in counseling research, we looked for direction to the field of clinical psychology, which has a growing subfield of implementation science that includes attention to treatment fidelity (Southam-Gerow & McLeod, 2013). One commonly cited approach in the development of treatment fidelity measures is the three-step method of Mowbray et al. (2003): (1) identifying and specifying fidelity criteria, (2) collecting data to measure fidelity to the identified model components, and (3) assessing the reliability and validity of the fidelity measure. We also referenced the Human Services Research Institute’s fidelity toolkit, which offers direction for the fidelity development process (Bond et al., 2000).

PURPOSE OF THE STUDY

In prior research, the authors have investigated the credibility and expectancy of nCBT with practitioners and clients. In this feasibility study into treatment fidelity, the authors sought to create and test an nCBT treatment fidelity checklist following the approach of Mowbray et al. (2003). In this article, we describe how we identified essential components of the model, developed an initial fidelity checklist, collected data to measure the identified model components, and conducted our initial examinations into adherence. In the

discussion section, we describe how the present study informed refinements to the treatment manual and fidelity measure. The following research questions guided the study: (1) What are the essential components of nCBT fidelity? (2) To what degree can expert raters determine when trained clinicians adhered to or deviated from the nCBT model using an initial nCBT treatment fidelity checklist?

Ultimately, we hoped that indications of IRR would establish a proof of concept for future use in a larger trial. We did not seek to address the third stage of fidelity development from Mowbray et al. (2003; i.e., reliability and validity testing), as this was a feasibility study with a smaller sample. We believed that the results of this feasibility study could have value to counselors who are seeking to research treatment fidelity or become trained in manualized treatments with fidelity stipulations.

METHODOLOGY

We addressed the two research questions by following the first two steps of the Mowbray et al. (2003) treatment fidelity development process. We defined the essential components of the model and created a checklist to measure the degree to which trained clinicians demonstrated these essential components of nCBT. We then used our initial treatment fidelity checklist to assess whether expert raters could determine when trained clinicians adhered to or deviated from the nCBT model. Expert raters evaluated 11 clinical sessions by nCBT-trained counselors.

Research Question 1: Defining Essential Components of nCBT Fidelity

To answer the first research question (What are the essential components of nCBT fidelity?), we followed the first step from Mowbray et al. (2003) of defining essential components of the model. We identified fidelity criteria and created a checklist to measure whether counselors were adhering to the essential components of the model required for treatment fidelity. At this stage, we sought to define content fidelity and process fidelity (Melnik & Morrison-Beedy, 2012). According to Mowbray et al. (2003), the development of fidelity criteria can be informed by three sources of information: an established model with proven efficacy, subject matter experts, and the opinions of clients regarding what seems to be effective about the model. When defining essential components, we sought to be informed by these three sources of information.

First, nCBT was developed as a revision of the well-established and highly researched CBT approach (see Field et al., 2015). When writing the treatment manual, we reviewed existing examples of treatment manuals and information about required components of the model (e.g., Champion & Power, 2012). As nCBT was already conceptualized to have three phases, we particularly referenced manuals with three treatment phases.

Second, we considered ourselves to be a panel of subject matter experts (Mowbray et al., 2003). We were the original creators of nCBT, had published the first article about the model (Field et al., 2015), had conducted numerous

nCBT trainings, and had previously conducted the only two existing empirical studies into the model (Field et al., 2016; Field et al., 2017). We met multiple times via a digital communication platform over the course of 6 months to write the 135-page treatment manual and craft the initial draft of the fidelity measure. During that process, we delineated the essential components of the model through collaborative decisions. We resolved discrepancies and conflict through consensus.

Third, prior to this study we had conducted a preliminary 12-month pilot study into the use of nCBT in a naturalistic setting with practitioners and clients (i.e., Field et al., 2016; Field et al., 2017). Those studies examined the counselor and client's evaluations of feasibility, utility, credibility, and expectancy of improvement when using the model. The authors used findings from those studies to inform the creation of a treatment manual, the delineation of essential model components from the treatment manual regarding both content and process fidelity, and the development of a 3-day training into the use of the treatment manual.

We prioritized the importance of adhering to process fidelity expectations when delineating the essential components of the model. To demonstrate fidelity to the nCBT model, trainees were required to follow the sequential phasic progression of establishing rapport and assessing physiological activation (Phase 1: attending to physiological reactions) before addressing Wave1 automatic responses (Phase 2: building the brain from the bottom up) and eventually Wave2 meta-cognitive appraisals (Phase 3: connecting the bottom to the top). We also identified essential components of content fidelity, believing that some form of neuroeducation (R. Miller, 2016) on Wave1 and Wave2 processing was an essential intervention of the model. Approved interventions were given as options for each phase, and clinicians had the freedom to choose among those specified interventions to individualize treatment. We considered the intervention options in each phase to have equivalent centrality to nCBT. In accordance with Melnyk and Morrison-Beedy (2012), we believed that adhering to process fidelity (i.e., treatment sequence) was a higher priority than adhering to specific content fidelity.

Once we identified essential components of the model, we completed the second step from Mowbray et al. (2003) of creating a treatment fidelity measurement tool. As recommended by the Human Services Research Institute (Bond et al., 2000), we met on multiple occasions via a digital communication platform to refine the items listed on the initial draft of the fidelity checklist. We reviewed traditional CBT fidelity scales (e.g., Muse, McManus, & Rakovshik, 2017) to inform how items could be operationalized and measured. We created a checklist for measuring both skill and fidelity, consistent with adherence measures in the literature (e.g., Interpersonal Psychotherapy Adherence and Quality Scale; Interpersonal Psychotherapy Institute, 2014).

We developed a first draft of the measure, entitled the nCBT Treatment Fidelity Checklist (nCBT-TFC). We used a 3-point Likert-type scale to assess

skill (0 = no evidence/absent, 1 = introductory level skill, 2 = advanced level skill, N/A = not observed) and fidelity (0 = does not follow principles, 1 = follows principles with deviation, 2 = follows principles without deviation, N/A = not observed) at each of the three treatment phases. This structure resulted in six data points that we used to evaluate interrater consistency: Phase 1 skill, Phase 1 fidelity, Phase 2 skill, Phase 2 fidelity, Phase 3 skill, and Phase 3 fidelity. See Appendix A for a review of the essential components as outlined in the fidelity checklist.

Research Question 2: Evaluating Adherence to the nCBT Model

After identifying the essential components of the model and creating an instrument to measure the essential components, we selected a descriptive quantitative design to evaluate the degree to which expert raters could determine whether trained clinicians adhered to or deviated from the nCBT model using the nCBT-TFC. In scoring adherence, we used IRR coefficients to evaluate rater consistency.

We recruited licensed clinicians for the study. Following institutional review board approval, we sent e-mail invitations to licensed counselors who had completed a 3-day face-to-face nCBT training (6 hours per day for 3 consecutive days). Details about the original training and associated study, including recruitment of participants, can be found in R. Miller, Field, Beeson, Doumas, and Jones (2018). We sent the first recruitment e-mail 6 weeks post-training to all attendees and the second recruitment e-mail 8 weeks post-training. We determined that a period of at least 6 weeks would be necessary because we believed that counselors needed time to review the treatment manual and practice nCBT in their current work settings. Our decision was not grounded in prior studies, as very little information has been published regarding observational coding methodology for fidelity of CBT interventions (Rodriguez-Quintana & Lewis, 2018).

After the initial recruitment, participants scheduled a time to record a mock nCBT session. For the recorded session, participants met with graduate student volunteers who role-played a simulated client presentation in a confidential university training clinic with built-in recording capabilities. Sessions lasted approximately 50 minutes. We asked participants to apply nCBT principles to the best of their abilities, while also approaching the therapeutic relationship in a natural and authentic manner consistent with how participants typically approached their first sessions. We gave participants introductory information about the mock client ahead of the recorded session. The information reflected a composite client from our clinical experience. In the introductory information, we indicated that the client was referred by their primary care physician after seeking medical attention for a panic attack they experienced at work.

Seven volunteer counseling students served as mock clients for the video-recorded session. A member of our research team gave each mock client the same script that detailed client characteristics, basic background informa-

tion, and current symptomology. This research team member then discussed and role-played the case in depth with each mock client to ensure that they demonstrated consistent in-session presentations.

Fidelity evaluation participants. A total of 11 participants (73% female, 27% male) responded to the recruitment e-mails and consented to participate in the study. Ages ranged from 33 to 53 ($M = 42.9$, $SD = 7.1$). The majority ($n = 8$) identified as White. One participant identified as White and Latino/a, and one participant identified as White and Native American. One participant identified as Black or African American. The majority of participants (91%, $n = 10$) held a professional counselor license, and one participant held a social work license. Years in clinical practice ranged from 1 to 10 ($M = 6$, $SD = 5$). Participants were given unique participant identification numbers to protect privacy.

Treatment evaluation measures. To assess nCBT treatment fidelity, each study author independently rated all 11 50-minute video-recorded sessions using the nCBT-TFC. After all research team members completed coding, each member independently entered their scores into a shared database for analysis. We also recorded individual memos when coding videos. During coding team meetings, we discussed key themes from memos, such as similarities and differences in how each coder interpreted data. Reviewing memos deepened discussions about refinements to essential model components and fidelity measurement.

Statistical analysis. We selected kappa correlation coefficients to evaluate IRR for multiple raters (McHugh, 2012). We followed Light's (1971) recommendation to use a corrected form of kappa coefficient for three or more raters, by computing the arithmetic mean of kappa coefficients (Hallgren, 2012). We also computed a Pearson product moment correlation coefficient to determine the relationship between participant skill and fidelity ratings.

Power analysis. The sample size met the 80% power criterion (30 observations per cell) required for analysis of variance comparisons by rater and by phase. For rater comparisons (four groups), we analyzed differences in mean ratings for 11 video observations across three phases with regard to both skill (33 observations per cell) and fidelity (33 observations per cell). For phase comparisons (three groups), we analyzed differences in mean ratings for 11 video observations across four raters (44 observations per cell). We were unable to analyze differences across student volunteers (seven groups), as this group variable had only 12 observations per cell for skill and fidelity (three ratings by four raters). We used 67% agreement and Cohen's kappa of .67 as a cutoff for acceptable IRR (Hallgren, 2012). According to Bujang and Baharum (2017), a 3-point rating scale requires a sample size of at least 10 for 80% power at the .05 alpha level.

Throughout the remainder of the manuscript, we refer to each individual rater as R1, R2, R3, or R4. Each number corresponds with authorship order (e.g., R1 pertains to the first author of this manuscript).

RESULTS

We found the nCBT-TFC a feasible though imperfect measure of whether trained clinicians adhered to or deviated from the nCBT model. We compared interrater consistency by each expert rater and by treatment phase.

Raters

IRR kappa coefficients indicated that our ratings were fairly consistent for a pilot instrument. The overall kappa coefficient for combined skill and fidelity ratings across all 11 videos was $\kappa = .73$ and ranged from $\kappa = .64$ to $\kappa = .84$. Kappa coefficients equaled or exceeded .67 for all but one participant, and as such, the nCBT-TFC met the criterion for adequate IRR for 10 out of the 11 participants.

We observed important differences in mean skill and fidelity ratings when raters were compared with each other. Table 1 depicts differences in mean overall ratings for R1 ($M = 0.81$, $SD = 0.47$), R2 ($M = 0.98$, $SD = 0.38$), R3 ($M = 0.70$, $SD = 0.53$), and R4 ($M = 0.77$, $SD = 0.44$). A one-way analysis of variance for these differences was significant, $F(3, 260) = 4.45$, $p < .05$, $\eta^2 = 0.05$, constituting a small- to medium-sized effect. Tukey HSD post hoc analyses revealed significant differences (all $ps < .05$) between R2 vs. R4 and R2 vs. R3. Therefore, it appears that raters had a small- to medium-sized impact on skill and fidelity ratings.

Skill and Fidelity

In addition, kappa coefficients for skill and fidelity ratings exceeded the .67 threshold for all treatment phases except Phase 3 skill (Phase 1 skill = .68, Phase 2 skill = .84, Phase 3 skill = .64; Phase 1 fidelity = .80, Phase 2 fidelity = .77, Phase 3 fidelity = .68). When averaged across treatment phases, individual skill and fidelity ratings were strongly correlated ($r = .79$, $p < .05$), representing a large-sized effect (Cohen, 1988). Skill and fidelity, therefore, may be potentially overlapping constructs that are difficult to differentiate. Consistent with this finding, we struggled to differentiate between skill and fidelity when coding videos.

Phases

Although skill and fidelity ratings appeared related to one another within phases, we also observed that they differed between phases. We gave the highest skill and fidelity ratings when assessing Phase 1 (skill $M = 1.02$, skill $SD = .37$; fidelity $M = 1.01$, fidelity $SD = .35$), the second highest ratings when assessing Phase 2 (skill $M = .78$, skill $SD = .40$; fidelity $M = .85$, fidelity $SD = .46$), and the lowest ratings when assessing Phase 3 (skill $M = .68$, skill $SD = .55$; fidelity $M = .55$, fidelity $SD = .48$). A repeated measures ANOVA was statistically significant for differences in skill, $F(2, 129) = 6.53$, $p < .05$, $\eta^2 = 0.09$, and fidelity, $F(2, 129) = 12.74$, $p < 0.05$, $\eta^2 = 0.16$, among phases, both representing medium- to large-sized effects (Cohen, 1988). Tukey HSD post hoc analyses revealed significant differences ($p < .05$) between Phases 1 and 3 and between

Table 1 Skill and Fidelity Ratings by Researcher (*M, SD*)

Researcher/coder	S1	S2	S3	F1	F2	F3	Overall (rater)
R1	1.05 (.35)	0.82 (.40)	0.68 (.51)	1.00 (.39)	0.77 (.52)	0.55 (.52)	0.81 (.47)
R2	0.95 (.37)	0.84 (.38)	1.05 (.40)	1.09 (.30)	1.07 (.42)	0.89 (.44)	0.98 (.38)
R3	1.14 (.45)	0.68 (.46)	0.27 (.41)	1.00 (.39)	0.73 (.47)	0.36 (.45)	0.70 (.53)
R4	0.93 (.32)	0.77 (.41)	0.73 (.61)	0.95 (.35)	0.82 (.40)	0.41 (.38)	0.77 (.44)
Overall (phase)	1.02 (.37)	0.79 (.40)	0.68 (.55)	1.01 (.35)	0.85 (.46)	0.55 (.48)	0.81 (.47)

Note. S1 = skill in Phase 1; S2 = skill in Phase 2; S3 = skill in Phase 3; F1 = fidelity in Phase 1; F2 = fidelity in Phase 2; F3 = fidelity in Phase 3; R1 = first-author researcher/coder; R2 = second-author researcher/coder; R3 = third-author researcher/coder; R4 = fourth-author researcher/coder.

Phases 2 and 3. Thus, treatment phase had a medium- to large-sized effect on skill and fidelity ratings.

These differences in ratings by phase matched our observations. We noticed that participants rarely seemed to follow the phasic sequential treatment progression and thus did not adhere to process fidelity. Instead, counselors moved too quickly into Phase 2 and 3 interventions, such as neuro-education on Wave1 and Wave2 processing, problem-solving, or facilitating guided exercises such as mental imagery, before they had adequately demonstrated Phase 1 skills such as active listening and assessing the client's physiological responding.

DISCUSSION

Through following the process outlined by Mowbray et al. (2003), we sequentially addressed each of the research questions guiding this feasibility study. We first identified the essential components of nCBT. Our first step was informed by reviews of established models with proven efficacy, using subject matter experts and reviewing prior nCBT studies for the opinions of counselors and clients regarding the model's utility, credibility, and expectancy of change occurring. We then drafted an instrument for measuring the essential components and conducted a study to measure the extent to which expert raters could determine when trained clinicians adhered to or deviated from the nCBT model using the initial nCBT-TFC instrument.

The nCBT-TFC approached adequate IRR, having met the .67 cutoff for acceptable kappa coefficients across all but one of the 11 participants. Although our overall average skill and fidelity ratings were consistent among raters, some variation did exist. Raters had a small- to medium-sized impact on skill and fidelity ratings, and some raters consistently gave higher ratings than others when viewing the same videos. This finding indicates that IRR could be further enhanced. Our discovery initiated further efforts to refine the essential components of the model at each phase and revise the treatment fidelity checklist to better represent the essential components.

Redefining Essential Components

We followed a recursive process of further clarifying essential components of the nCBT model given findings from the present study. The treatment phase had a medium- to large-sized impact on skill and fidelity ratings. Despite the research team originally defining treatment fidelity as following the phasic sequential progression of nCBT, participants still moved too quickly into problem-solving potential solutions, and their approach seemed too directive and potentially invalidating to clients. The nCBT-TFC was modified to further clarify and emphasize the importance of following the phasic sequence of client assessment and attending to physiological responding prior to initiating Wave1 or Wave2 interventions. We also modified the treatment manual and 3-day training materials to emphasize the requirement for counselors to assess and attend to physiological responding first, in the hope of reducing therapist drift.

Our recognition that certain essential elements had to be present for nCBT interventions to be effective appears to mirror evolutions within other major therapeutic approaches. For example, in discussing the evolution of motivational interviewing (MI), W. R. Miller and Rollnick (2009) referred to “the spirit of MI” (e.g., collaborative approach, emphasizing client autonomy, evoking clients’ own reasons for change). W. R. Miller and Rollnick (2009) noted “the practice of MI without understanding and manifesting this spirit is like the words without the music of a song” (p. 131). We similarly found that when participants applied components of nCBT without the use of basic microskills or without apparent intentionality in delivery (e.g., following nCBT sequence), they were not actually practicing effective nCBT.

Refining Measures

Study findings indicated that the fidelity checklist needed further operationalization. We theorized that inconsistencies among raters could have resulted from the fidelity checklist being too broad, open to interpretation, and lacking adequate definition. The large-sized correlation ($r = .79$) between skill and fidelity suggested that the nCBT-TFS may in fact be measuring a single construct. This surprised us, as we observed some participants demonstrating skills but not fidelity, and vice versa. To address the probable lack of differentiation between skill and fidelity, the checklist was revised using the constructs of structure, process, and delivery in the extant literature (Melnik & Morrison-Beedy, 2012). We restructured the rating rubric for each phase to include the requirement of applying the attend–build–connect (A-B-C) progression both across multiple sessions (i.e., treatment structure) and within the same session (i.e., treatment process). For example, the counselor should first assess the client and attend to physiological reactions before providing Wave1 or Wave2 interventions, and subsequently help the client process their experience of the intervention to reinforce learning and self-efficacy. We also included information about how fidelity should be measured. We required that formal documents such as treatment plans and session summaries be used to assess

intentional attempts at using the attend–build–connect progression across multiple sessions, but that video recordings be used to evaluate whether the attend–build–connect progression was evident within the same session. This revision reflected our observation that a single video recording was insufficient to fully evaluate treatment fidelity. The revised nCBT-TFC is included in Appendix B.

Implications for Assessing Treatment Fidelity

We propose several considerations for counselor researchers and practitioners regarding the assessment of the treatment fidelity, grounded in the findings from the study.

Training components. Trainings in manualized interventions may benefit from several components that could enhance fidelity. Assessing basic attending skills during training could help to identify trainees who need additional training to acquire the basic counseling skills necessary for treatment fidelity. Demonstration videos following a training may be helpful for reinforcing how to implement the treatment approach. Ongoing post-training supervision and consultation could help trainees to continue developing skills and fidelity in the method, particularly if the supervision or consultation involves observation through live supervision or review of video recordings. Experts from the National Institute of Health’s Behavior Change Consortium (Bell et al., 2004) conceptualized training as an ongoing process rather than a one-time training course, to protect against therapist drift. Beidas, Edmonds, Marcus, and Kendall (2012) similarly reported that post-training consultation enhances treatment fidelity.

Observations. The study design presented potential measurement problems regarding how to appraise nCBT fidelity. Multiple observations may be useful to adequately assess skill and fidelity across different interventions and phases. We analyzed a sole 50-minute video in this study, which may not have been an accurate reflection of counselor nCBT skill and fidelity across phases. Participants received higher skill and fidelity ratings for Phase 1 compared to Phases 2 and 3, which could have been influenced by the single-session nature of the simulated role-play. Because the simulated role-play was a one-time meeting, participants may have approached the simulation as an intake session. Although perhaps optimal, using multiple observations when assessing skill and fidelity is somewhat challenging for research teams given the substantial resources that would be required to manage the volume of data collection and analysis. We found that 11 hours of video was cumbersome to code, and coding multiple sessions would require a fairly large research group and/or a larger amount of time for coding.

Requiring both content and process fidelity. Creating a fidelity checklist requires a balance that honors the individualized, unfolding nature of counseling encounters, while at the same time providing some standardized guidelines for the approach. A balance has to be struck between flexibility and prescriptive rigidity in how counselors are directed to use a treatment manual. Ray et al. (2017) described a similar struggle, commenting on the “inherent philosoph-

ical inconsistencies between [child-centered play therapy] and procedures” (p. 214).

The fidelity checklist used in this study only required counselors to follow a phasic progression and did not require specific skills to be used (i.e., required process fidelity and not content fidelity; Melnyk & Morrison-Beedy, 2012). The flexibility of this approach naturally limited treatment fidelity, evident in the variation and inconsistency we observed in participant videos. We decided at the conclusion of this study to introduce more clarity and standardized guidelines for counselors demonstrating the essential components of the model.

Limitations

Our study had several limitations. The size of the participant sample was small ($N = 11$), and all statistical findings must be interpreted with caution. Although the sample size was acceptable for a feasibility study, a larger sample is needed to complete the third step from Mowbray et al. (2003) of psychometrically evaluating the reliability and validity of the measure. This study resulted in revisions to the original nCBT-TFC, and further studies are needed to appraise the reliability and validity of the revised checklist. For example, a subsequent larger scale study could evaluate whether checklist scores actually predict adherence.

It is possible that the simulated nature of the study influenced counselor behaviors during the recorded session with the mock client. The staged scenario may not be an accurate simulation of how counselors would respond when working with an actual client. Future studies could review video recordings of counselors providing services to actual clients. The review of 50-minute videos was cumbersome, and a larger scale study should consider how to better assess treatment fidelity across sessions without coding multiple hours of video for each trainee. In addition, further studies are needed to evaluate the impact of variables such as age, race/ethnicity, sex, and years of experience. These demographic variables might have influenced treatment fidelity ratings in this study. Finally, our research team consisted of the developers of nCBT, and the results may have been biased by researcher allegiance. Future studies will need to be conducted by researchers without ties to the developers of nCBT.

Summary and Recommendations

This feasibility study described the process of moving through the steps outlined by Mowbray et al. (2003). This study represents a starting point for psychometrically evaluating the nCBT-TFC. Both the development of a new clinical approach and the effort to develop a fidelity checklist to assess that approach are uncommon endeavors to emerge from the counseling field specifically. From this study, we have several recommendations for counselors who wish to define and measure treatment fidelity, as well as for counselors who wish to learn and demonstrate fidelity to a manualized treatment.

We have four recommendations for counselor researchers seeking to define and measure treatment fidelity. First, working in teams with the origi-

nators of the approach is useful when making decisions about essential model components and how best to measure them, as the originators of the approach can serve as subject matter experts. Second, identifying essential components is a complex task that requires thought into which components make the model unique and are central to its identity. Third, collecting and analyzing observational data on treatment fidelity measures can be cumbersome and time intensive, and may require substantial research support. Fourth, multiple sources of data (e.g., documents, multiple recordings) may be required to adequately measure fidelity throughout the treatment process. To meet these recommendations, thought must be given to the process of evaluating a trainee across multiple sessions without coding several hours of video recordings for each trainee. This is especially important for large-scale studies. We hope that these recommendations will assist counseling researchers who are seeking to measure treatment fidelity for new counseling theories.

For practitioners, we recommend fully developing and/or maintaining intentional practice of basic counseling skills prior to learning and implementing a manualized treatment, as deficits in these skill areas appear likely to impact treatment fidelity. This step likely requires counselors to record and watch themselves (and/or have a peer watch the recordings) to ensure they are truly demonstrating adequate use of microskills. In this study, we found a strong correlational relationship between ratings of counselor skill and treatment fidelity ($r = .79, p < .05$). The authors observed that counselors who were weaker in basic counseling skills such as attending and summarizing were less likely to demonstrate fidelity when providing more advanced techniques such as psychoeducation on the Wave model. We found it surprising that some participants struggled with these basic skills, considering that participants were licensed mental health professionals. A key finding from this study is therefore the ongoing need for training and supervision in basic counseling skills for some practitioners, even post-licensure.

REFERENCES

- Barsaglini, A., Sartori, G., Benetti, S., Petterson-Yeo, W., & Mechelli, A. (2014). The effects of psychotherapy on brain function: A systematic and critical review. *Progress in Neurobiology*, *114*, 1–14. <https://doi.org/10.1016/j.pneurobio.2013.10.006>
- Beck, A. T., & Haigh, E. A. P. (2014). Advances in cognitive theory and therapy: The generic cognitive model. *Annual Review of Clinical Psychology*, *10*, 1–24. <https://doi.org/10.1146/annurev-clinpsy-032813-153734>
- Beeson, E. T., Field, T. A., Jones, L. K., & Miller, R. (2017). *nCBT: A semi-structured treatment manual*. Boise, ID: Boise State University.
- Beidas, R. S., Edmunds, J. M., Marcus, S. C., & Kendall, P. C. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services*, *63*, 660–665. <https://doi.org/10.1176/appi.ps.201100401>
- Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Mimicucci, D. S., Ory, M., ... Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, *23*, 443–451. <https://doi.org/10.1037/0278-6133.23.5.443>
- Blom, E. H., Duncan, L. G., Ho, T. C., Connolly, C. G., LeWinn, K. Z., Chesney, M., ... Yang, T. T. (2014). The development of an RDoC-based treatment program for adolescent

- depression: “Training for Awareness, Resilience, and Action” (TARA). *Frontiers in Human Neuroscience*, 8, 1–19. <https://doi.org/10.3389/fnhum.2014.00630>
- Blom, E. H., Tymofflyeva, O., Chesney, M. A., Ho, T. C., Moran, P., Connolly, C. G., ... Yang, T. Y. (2017). Feasibility and preliminary efficacy of a novel-RDoC-based treatment program for adolescent depression: “Training for Awareness Resilience and Action” (TARA)—a pilot study. *Frontiers in Psychiatry*, 7, 1–13. <https://doi.org/10.3389/fpsy.2016.00208>
- Bond, G., Williams, J., Evans, L., Salyers, M., Kim, H. W., & Sharpe, H. (2000). *Psychiatric rehabilitation fidelity tool*. Cambridge, MA: Human Services Research Institute.
- Bujang, M. A., & Baharum, N. (2017). Guidelines of the minimum sample size requirements for Cohen’s kappa. *Epidemiology, Biostatistics, and Public Health*, 14(2), 1–10. <https://doi.org/10.2427/12267>
- Champion, L., & Power, M. J. (2012). Interpersonal psychotherapy for eating disorders. *Clinical Psychology and Psychotherapy*, 19(2), 150–158. <https://doi.org/10.1002/cpp.1780>
- Clark, D. A., & Beck, A. T. (2010). Cognitive theory and therapy of anxiety and depression: Convergence with neurobiological findings. *Trends in Cognitive Sciences*, 14, 418–424. <https://doi.org/10.1016/j.tics.2010.06.007>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.
- David, D., Cristea, I., & Hofmann, S. G. (2018). Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Frontiers in Psychiatry*, 9(4), 1–3. <https://doi.org/10.3389/fpsy.2018.00004>
- Field, T. A., Beeson, E. T., & Jones, L. K. (2015). The new ABCs: A practitioner’s guide to neuroscience-informed cognitive-behavioral therapy. *Journal of Mental Health Counseling*, 37, 206–220. <https://doi.org/10.17744/mehc.37.3.02>
- Field, T. A., Beeson, E. T., & Jones, L. K. (2016). Neuroscience-informed cognitive behavior therapy in clinical practice: A preliminary study. *Journal of Mental Health Counseling*, 38, 139–154. <https://doi.org/10.17744/mehc.38.2.05>
- Field, T. A., Beeson, E. T., Jones, L. K., & Miller, R. (2017). Counselor allegiance and client expectancy in neuroscience-informed cognitive-behavior therapy: A 12-month qualitative follow-up. *Journal of Mental Health Counseling*, 39, 351–365. <https://doi.org/10.17744/mehc.39.4.06>
- Gonçalves, Ó. F., & Perrone-McGovern, K. M. (2016). Translating neuroscience into counselling practice. *Canadian Journal of Counselling and Psychotherapy*, 50, 421–440. Retrieved from <https://cjc-rcc.ualgary.ca/article/view/61122>
- Gros, D. F. (2015). Design challenges in transdiagnostic psychotherapy research: Comparing Transdiagnostic Behavior Therapy (TBT) to existing evidence-based psychotherapy in veterans with affective disorders. *Contemporary Clinical Trends*, 43, 114–119. <https://doi.org/10.1016/j.cct.2015.05.01>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8, 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Interpersonal Psychotherapy Institute. (2014). *IPT adherence and quality scale*. Retrieved from <http://iptinstitute.wpengine.com/wp-content/uploads/2015/09/IPT-Quality-Adherence-Scale.pdf>
- Light, R. J. (1971). Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological Bulletin*, 76, 365–377. <https://doi.org/10.1037/h0031643>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemical Medicine*, 22, 276–282. <https://doi.org/10.11613/BM.2012.031>
- McRae, K., Misra, S., Prasad, A. K., Periera, S. C., & Gross, J. J. (2012). Bottom-up and top-down emotion generation: Implications for emotion regulation. *Social Cognitive and Affective Neuroscience*, 7, 253–262. <https://doi.org/10.1093/scan/nsq103>
- Melnyk, B. M., & Morrison-Beedy, D. (2012). *Intervention research: Designing, conducting, analyzing, and funding*. New York, NY: Springer.
- Miller, R. (2016). Neuroeducation: Integrating brain-based psychoeducation into clinical practice. *Journal of Mental Health Counseling*, 38, 103–115. <https://doi.org/10.17744/mehc.38.2.02>
- Miller, R., Field, T. A., Beeson, E. T., Doumas, D. M., & Jones, L. K. (2018). *The impact of neuroscience-informed cognitive-behavior therapy training on knowledge and interoceptive awareness*. Manuscript submitted for publication.

- Miller, W. R., & Rollnick, S. (2009). Ten things that motivational interviewing is not. *Behavioural and Cognitive Psychotherapy*, 37, 129–140. <https://doi.org/10.1017/S1352465809005128>
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24, 315–340. <https://doi.org/10.1177/109821400302400303>
- Muse, K., McManus, S., & Rakovshik, S. (2017). Development and psychometric evaluation of the Assessment of Core CBT Skills (ACCS): An observation based tool for assessing Cognitive Behavioural Therapy competence. *Psychological Assessment*, 29, 542–555. <https://doi.org/10.1037/pas0000372>
- Ochsner, K. N., Ray, R. R., Hughes, B., McRae, K., Cooper, J. C., Weber, J., ... Gross, J. J. (2009). Bottom-up and top-down processes in emotion generation: Common and distinct neural mechanisms. *Psychological Science*, 20, 1322–1331. <https://doi.org/10.1111/j.1467-9280.2009.02459.x>
- Onken, L. S., Carroll, K. M., Shoham, V., Cuthbert, B. N., & Riddle, M. (2014). Reenvisioning clinical science: Unifying the discipline to improve the public health. *Clinical Psychological Science*, 2, 22–34. <https://doi.org/10.1177/2167702613497932>
- Ray, D. C., Purswell, K., Haas, S., & Aldrete, C. (2017). Child-centered play therapy-research integrity checklist: Development, reliability, and use. *International Journal of Play Therapy*, 26, 207–217. <https://doi.org/10.1037/pla0000046>
- Reiser, R. P., & Milne, D. L. (2014). A systematic review and reformulation of outcomes evaluation in clinical supervision: Applying the fidelity framework. *Training and Education in Professional Psychology*, 8, 149–157. <https://doi.org/10.1037/tep0000031>
- Rodriguez-Quintana, N., & Lewis, C. C. (2018). Observational coding training methods for CBT treatment fidelity: A systematic review. *Cognitive Therapy and Research*, 42, 358–368. <https://doi.org/10.1007/s10608-018-9898-5>
- Southam-Gerow, M. A., & McLeod, B. D. (2013). Advances in applying treatment integrity research for dissemination and implementation science: Introduction to special issue. *Clinical Psychology: Science and Practice*, 20, 1–13. <https://doi.org/10.1111/cpsp.12019>
- Waller, G., & Turner, H. (2016). Therapist drift redux: Why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Research and Therapy*, 77, 129–137. <https://doi.org/10.1016/j.brat.2015.12.005>

APPENDIX A: FIRST DRAFT OF NEUROSCIENCE-INFORMED COGNITIVE-BEHAVIOR TREATMENT FIDELITY CHECKLIST

Skill scale: 0 = No evidence/absent, 1 = introductory-level skill, 2 = advanced-level skill (N/A = not observed)

Fidelity scale: 0 = Does not follow principles, 1 = Follows principles with deviation, 2 = Follows principles without deviation (N/A = not observed)

A (Attend to Physiological Reactions)

Counselor assesses physiological reactions, predominant response, treatment expectancy, and lifestyle habits that support optimal brain health and prevent neurophysiological vulnerability.

Skill: _____ **Fidelity:** _____

- Conducting a thorough neurobiopsychosocial assessment
- Develops rapport and trust with client through basic microskills (active listening, reflection, summarizing) and core conditions (empathic skills, warmth, concern, positive regard, and genuineness)
- Providing psychoeducation on Wave 1 and 2

- In-session observing of physiological response, and asking client if they have also noticed this response
- Appraising out-of-session physiological reactions
- Assessing predominate response
- Inquiring about client motivation and expectancy for change to occur
- Providing appropriate referrals for adjunctive or alternative treatments
- Reviewing “Supporting Optimal Brain Health and Preventing Neurophysiological Vulnerability” handout
- Collaboratively completing the “Exploration of Current and Future Activities” worksheet
- Reviewing client weekly responses on “Activity Tracking Sheet” in subsequent sessions

B (Build the Brain from the Bottom-up)

Counselor and client collaboratively identify practices that assist with achieving preconscious emotional and physiological self-regulation (Wave1 strategies).

Skill: _____ **Fidelity:** _____

- Affective modeling and interactive regulation
- Collaboratively identifying sensory-based coping strategies (auditory, visual, olfactory, gustatory, tactile), to assist with grounding and coping during physiological arousal, using established guidelines for selecting sensory-based coping strategies
- Collaboratively identifying daily mindfulness practice
- Conditioned anchoring
- Collaboratively defining exposure hierarchy, choosing the first exposure task, and assessing readiness
- Biofeedback and neurofeedback
- Identifying real-life situations for safely practicing daily exercises that facilitate state-dependent learning
- Inquiring about outside-of-session practice at each session

C (Connect the Bottom to the Top)

Counselor assists client to become aware and attuned of their responding, to connect behaviors and emotions to their physiological state, to develop self-acceptance and compassion, and to re-appraise past events and explore implicit schema that generate further emotional, physiological, and behavioral consequences (Wave2 strategies).

Skill: _____ **Fidelity:** _____

- Assists client to self-monitor through providing feedback about physiological responding
- Processing with client, “if your body could talk, what would it say?”
- Counting pulse exercise
- Uses “Connecting Behaviors and Emotions to Physiological States” worksheet

- Uses sensory-based exploration
- Body scan
- Self-compassion meditation: 1. “Sit with” an experience rather than approach (move toward) or avoid (move away), 2. Words of gratitude, understanding, and compassion for different parts of your body?
- Reappraisal, using experience as disputation
- Using thought records
- Explores implicit schema through downward arrow and/or free association exercises. Follows directions and protocol when using these techniques. Identifies deeper meaning/values, and helps client develop goals to live out these values.

APPENDIX B: SECOND DRAFT OF NEUROSCIENCE-INFORMED COGNITIVE-BEHAVIOR THERAPY TREATMENT FIDELITY CHECKLIST

	Absent or minimal (0)	Partial (1)	Full (2)
Structure	Does not document the attend–build–connect sequence across multiple sessions.	Treatment plan and session summaries partially document the attend–build–connect progression across multiple sessions: <ul style="list-style-type: none"> • First determines predominant response process and style • Selects Wave 1 and 2 interventions without consideration for earlier assessment (Note: Determining the predominant response process and style is required for partial adherence to the model. Selecting Wave 1 or 2 interventions without determining the predominant response process and style does not count as partial adherence to the model)	Treatment plan and session summaries document the attend–build–connect progression across multiple sessions: <ul style="list-style-type: none"> • First determines predominant response process and style • Selects Wave 1 and 2 interventions accordingly

<p>Process</p>	<p>Does not demonstrate the attend–build–connect sequence within same session.</p>	<p>Sections of video recording partially demonstrate the attend–build–connect progression within a single session:</p> <p>Attending skills</p> <ul style="list-style-type: none"> • Attends to physiological reactions, shares observations with client, helps client “sit with” physiological reactions • Provides education on Wave 1 and 2 <p>(Note: Attending to physiology skills are required for partial adherence to the model. Using Wave 1 or 2 interventions in an appropriate manner without first demonstrating attending to physiology skills does not count as partial adherence to the model)</p>	<p>Sections of video recording demonstrate the attend–build–connect progression within a single session:</p> <p>Attending skills</p> <ul style="list-style-type: none"> • Attends to physiological reactions, shares observations with client, helps client “sit with” physiological reactions • Provides education on Wave 1 and 2 <p>Building skills</p> <ul style="list-style-type: none"> • Models affect regulation • Collaboratively plans (1) parasympathetic response in non-threat situations and (2) state-dependent learning in stressful situations • Coping strategies are portable, accessible, acceptable, and practiced repeatedly • Reduces susceptibility through lifestyle change • Uses appropriate bottom-up techniques within scope of training <p>Connecting skills</p> <ul style="list-style-type: none"> • Self-acceptance and compassion • Using experience to dispute • Uses appropriate top-down techniques within scope of training
<p>Delivery</p>	<p>Sections of video recording demonstrate:</p> <ul style="list-style-type: none"> • Inappropriate content for level of comprehension • Inappropriate nonverbals such as vocal tone and pitch • Judgmental or defensive • Inaccurate reflection and summarization of verbal content 	<p>Sections of video recording demonstrate:</p> <ul style="list-style-type: none"> • Appropriate content for client’s level of comprehension • Appropriate nonverbals such as vocal tone/pitch • Nonjudgmental and nondefensive • Accurate reflection and summarization of verbal content • Mistimed responses, or inappropriate phrasing or pacing • Includes unnecessary details • Confidence is not always demonstrated 	<p>Sections of video recording demonstrate:</p> <ul style="list-style-type: none"> • Appropriate content for client’s level of comprehension • Appropriate nonverbals such as vocal tone/pitch • Nonjudgmental and nondefensive • Accurate reflection and summarization of verbal content • Appropriate timing, phrasing, pacing • Excludes unnecessary details • Demonstrates confidence