Developing the Interrater Reliability of the Modified EMDR Fidelity Checklist

Rebecca Z. Cooper
Anita D. Smith
David Lewis
School of Psychology and Exercise Science, Murdoch University, Western Australia

Christopher W. Lee
School of Psychology and Exercise Science, Murdoch University, Western Australia
School of Psychiatry, University of Western Australia, Western Australia

Andrew M. Leeds
Sonoma Psychotherapy Training Institute, Santa Rosa, California

Although treatment fidelity measures for eye movement desensitization and reprocessing (EMDR) have been cited in past research, none have been subject to any empirical investigation of reliability. This three-phase study aimed to quantify the interrater reliability of a measure of EMDR treatment fidelity. First, two raters refined the reprocessing section of the EMDR Fidelity Checklist (Leeds, 2016) by developing a descriptive item-by-item scoring system to improve interpretation and reliability. The resultant checklist was piloted on recordings of five EMDR session recordings from the Laugharne et al. (2016) study. The checklist was then revised. Next, the raters used the checklist to assess 15 other recorded EMDR sessions from the same study. The intraclass correlations (ICCs) were in the excellent range for all subscales and total session scores (i.e., >0.75), with an exception of the Desensitization subscale, ICC = 0.69 (0.08, 0.90). Finally, individual items in that subscale were evaluated, finding that five items did not contribute to the ICC. When these were removed/revised, the ICC for this subscale moved into the excellent range, ICC = 0.81(0.43, 0.94). The findings of this study indicate that this checklist may be a reliable measure of treatment fidelity for single reprocessing EMDR sessions with the possible exception of the Body Scan phase. Future research using the checklist with raters who were not involved in checklist development is needed to confirm the generalizability of these findings.

Keywords: eye movement desensitization and reprocessing; treatment fidelity; interrater reliability; post-traumatic stress disorder; psychometric measurement

Treatment fidelity (also known as treatment integrity) has been a key concern of researchers and clinicians for some time, yet little research attention has been dedicated to its investigation. It is defined as the extent to which a treatment was administered as described (Gearing et al., 2011). Historically, the most reliable method of measuring this construct has been through the use of quantitative questionnaires (Borelli, 2011). It is applied in clinical settings to monitor and ensure the adequacy of clients' exposure to a psychological intervention before assessing their treatment outcomes (Gearing et al., 2011). In treatment efficacy and effectiveness research, fidelity assessment functions as a manipulation check to control the risk of type II error as well as a form of quality control for ensuring standardized administration of an intervention (Schoenwald, 2011). Doing so reduces unintended variability in treatment effect and supports external validity by allowing for replication (Borelli, 2011).

 http://dx.doi.org/10.1891/1933-3196.13.1.32

Although treatment fidelity strategies are rarely incorporated in studies of psychological interventions (Perepletchikova, Treat, & Kazdin, 2007), their utility has often been demonstrated for strengthening the outcomes of psychological interventions, particularly those that are structured, manualized, and technique-driven. Durlak and DuPre (2008) in their meta-analysis of over 500 structured treatment programs found that effect sizes were two to three times greater when programs were carefully implemented and free from treatment fidelity issues. In Waller's (2009) review of the cognitive-behavioral therapy (CBT) literature, he found that CBT studies can be vulnerable to therapist drift (i.e., the phenomenon of experienced therapy administrators deviating from their training protocol over time), arguing that this leads to less effective CBT implementation and recommended retraining, regular supervision, and fidelity measurement to address this issue. Fidelity measures have been developed and tested for various psychological therapies such as CBT (Bassett, Stein, Rossi, & Martin, 2016; Lu et al., 2012; Southam-Gerow et al., 2016), dialectic behavioral therapy (DBT) (McCay et al., 2016), and mindfulness-based cognitive therapy (MBCT) (Prowse, Meadows, & Enticott, 2015).

Fidelity measurement has become particularly pertinent to eye movement and desensitization and reprocessing (EMDR). In Maxfield and Hyer's (2002) meta-analysis, they investigated the relationship between the methodological strengths and weaknesses of EMDR studies and their outcomes, finding that the incorporation of treatment fidelity design features moderately predicted stronger treatment outcomes in adult PTSD populations. In addition, Lee and Cuijpers' (2014) meta-analysis investigating the incremental effect of horizontal eye movements on the processing of emotional memories found that the incorporation of treatment fidelity strategies moderately predicted stronger effect sizes. In addition, Jeffries and Davis (2012) commented on the lack of treatment fidelity strategies in their systematic review of randomized controlled trials assessing EMDR efficacy for adult PTSD populations.

As yet, there is a scarcity of measures for EMDR fidelity in recent research. In addition, no study, to date, has reported any reliability data on the scale that has been used. Van der Kolk et al. (2007) conducted a randomized-controlled trial comparing the efficacy of EMDR versus fluoxetine and pill placebo to treat PTSD while using an early version of the now-revised EMDR Fidelity Rating Scale (EFRS; Korn, Maxfield, Smyth, & Stickgold, 2001, 2017). Adler-Tapia and Settle (2009) used an earlier version of their own EMDR fidelity checklist for working with children and adolescents (Adler-Tapia & Settle, 2008, 2016) in their pilot study to assess the effect of EMDR on childhood depression.

Leeds (2016) provides a set of six fidelity checklists covering all eight phases of EMDR (in his Appendices A.1–A.6) as well as procedural scripts and forms for planning and documenting EMDR treatment (in his Appendix B). Appendix A.5 of his guide provides a checklist (hereafter referred to as the "reprocessing checklist") which covers a single reprocessing session's rating scale (i.e., phases 3–8 of EMDR). Items are drawn from Shapiro's (2001) EMDR protocol. Clinicians watch an EMDR reprocessing session and rate the extent to which the treating EMDR clinician completed certain actions as requested by each item from "0" (i.e., no adherence) to "2" (i.e., good adherence). It also includes optional and conditional items which cover specific procedural recommendations for particular client scenarios according to the EMDR manual (Shapiro, 2001). Scores are averaged at the end of each phase to obtain an average subscale fidelity score. For the purposes of this study, we also averaged subscale fidelity scores into an overall score, which represents the overall treatment fidelity of the reprocessing session.

When assessing fidelity using a checklist, interrater reliability becomes particularly pertinent. Interrater reliability is the extent to which two raters achieve similar evaluations when using the same measure to evaluate the same event (Gwet, 2014). As a construct, it provides confidence in the generalizability of a single rater's scores to other raters. This is not to be confused with validity, which is the extent to which an instrument reflects the target construct (in this case, treatment fidelity; Hallgren, 2012). These two concepts are related in that the achievement of a measure's validity is dependent on its reliability; however, reliability in itself is not a sufficient indicator of validity (Hallgren, 2012). In fact a measure may have excellent reliability, where the error variance of the scores has been controlled to a high degree, yet achieve poor validity due to difficulties generalizing these scores to real-world applications.

In this study, the interrater reliability of Leeds' (2016) EMDR Fidelity Checklist for a reprocessing session was assessed. This involved developing rating descriptions for each item to further assist with interpretation and scoring of each item (see Appendix A). Audiovisual recordings of EMDR sessions on adults with PTSD were then scored using the new scale. ICC analyses were used to measure the interrater reliability of this reprocessing checklist at both the subscale and overall level.

Method

Participants

Participation occurred as part of a larger study on the impact of EMDR and prolonged exposure (PE) on the volumes of the amygdala and the hippocampus (Laugharne et al., 2016). These clients had been recruited from various services in Perth, Western Australia. Eligibility requirements for the study were that the participant be aged between 18 and 65 years and satisfied *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition (*DSM-IV*) criteria for PTSD. Participants were excluded if they had received either EMDR or PE in the past or were currently receiving another trauma-focused intervention. Participants were also excluded if they met criteria for a psychotic illness or a cluster B personality disorder, or if they had a substance dependency disorder. The 10 clients who received EMDR were on average 39.70 years old (standard deviation [*SD*] = 9.55) and 8 were female. Further details of the participants and method for can be found in the Laugharne et al. (2016). All participants provided informed consent and the study was approved by the human ethics research committee at Murdoch University (approval number 2011/161).

Practitioners

Two clinicians administered EMDR to clients. These comprised a postgraduate therapist and a registered mental health nurse who were trained in EMDR therapy by a trainer approved by the EMDR International Association (EMDRIA). Before clients were recruited, practitioners demonstrated competency by presenting video recordings of their pilot administrations of EMDR to their peer-supervision group and site coordinator. These clinicians were provided with regular supervision during the course of EMDR administration to ensure treatment competence.

Raters

Two postgraduate psychology students (authors RC and AS) were used to assess the EMDR practitioners for treatment fidelity and thereby develop the reprocessing checklist's interrater reliability. These raters attended an accredited 3-day workshop given by an EMDRIA approved trainer. These raters assessed 20 recordings in total (5 during the pilot trial and 15 during the interrater reliability development of the reprocessing checklist) under the supervision of a clinical psychologist who is an accredited EMDRIA trainer (author CL) as well as a second supervisor (author DL).

Design

This study consisted of three phases: (1) development of the reprocessing checklist's scoring instructions, pilot testing of the five sessions, and subsequent revision of items; (2) ratings of the 15 sessions; (3) revising the Desensitization items to improve this subscale's ICC. In phase 1, the item-by-item scoring instructions were created and agreed upon between the authors of this study before being tested on five pilot EMDR recordings. After this, feedback was exchanged between the raters and supervisors regarding the scoring instructions and these were finalized. In phase 2, the reprocessing checklist (with the finalized scoring instructions) was tested on 15 EMDR recordings and the scores from these formed the bases of the findings of this study. In phase 3, the Desensitization subscale was found to require improvements in terms of its ICCs. To this end, several items were identified as underperforming and were either deleted or rescored.

To develop and assess the interrater reliability of the reprocessing section of the EMDR Fidelity Checklist, a correlational design was used. The two primary variables in this regard comprised of each rater's average fidelity scores for the recordings of all EMDR sessions on a subscale level as well as an overall level. Subscale level fidelity scores were averaged across all completed items within an EMDR phase and were scored for each rater and for each session. Likewise, overall fidelity scores comprised mean fidelity score averaged across all completed subscales per rater per session. These average scores were obtained from the reprocessing fidelity checklist on a subscale level as well as on an overall level (see materials section below for more detail). Using these scores, ICCs were calculated between the raters to determine subscale and overall interrater reliability for the checklist.

Materials

EMDR Therapy Fidelity Rating Scale for Reprocessing Session (Appendix A.5)—Modified. Modifications were made by the raters to the Leeds (2016) original reprocessing section of the Fidelity Checklist during the supervised pilot trial of the checklist to improve consistency in the interpretation of certain items. This involved raters assessing the pilot recordings both cooperatively and independently so that any disagreements in scoring methods could be identified. Where disagreements were observed, these were discussed with thesupervisor. Several items on the checklist were further edited to improve clarity. Following this, comprehensive item-by-item instructions describing the conditions under which a score may be achieved

were added to the checklist to standardize the scoring process between raters. The author of this checklist was consulted before the modified checklist was finalized. It was this version which was used by raters to independently assess the remaining 15 recordings for interrater reliability (see Appendix).

Not all of the scales from the Leeds original checklist were assessed. Only the A.5 scale was examined, which contains 45 items covering all the aspects of a reprocessing session. There are six subscales in the A.5 which correspond to the six EMDR phases that are most relevant to the reprocessing of stressful or traumatic memories. These subscales were Reevaluation (4 items), Assessment (10 items), Desensitization (14 items), Installation (6 items), Body Scan (4 items), and Closure (7 items). All items were rated on a 3-point scale (0 = "missing or no adherence," 1 = "adherence is identified but is weak or flawed," 2 = "adherence is good"). Scores for each subscale were averaged to obtain fidelity scores for each EMDR phase as well as for the overall EMDR session.

Completion of every subscale or item was not required to appropriately administer this checklist. Except for the Reevaluation and Closure subscales (which are both mandatory in the checklist), the initiation of each subscale was contingent upon the successful completion of the previous subscale, which is in accordance with EMDR practice (Shapiro, 2018). Similarly, many of the items of this checklist were optional or conditional upon a particular event occurring in-session (e.g., four items from the Desensitization subscale could be assessed only if processing of the target had become ineffective). This meant that not every item on this checklist was assessable while rating an EMDR recording. Hence, it was possible to obtain similar subscale and overall average scores for this measure through the completion of different items or subscales.

Procedures

Clients in the original study attended two 90-minute EMDR therapy sessions each week for 6 to 8 weeks. The course of EMDR ended either when the client's targets had been successfully reprocessed as per Shapiro's (2001) guidelines or when the client had completed 12 sessions. A selection pool of video recordings (i.e., two recordings per client) was created for rater assessment using a stratified randomized sampling technique. For each client, this involved selecting one video recording from the first six sessions of each client and then another from their remaining six sessions, which enabled the investigators to control the effects of therapist drift. Intake sessions involving the first two phases of EMDR therapy were excluded from the sampling procedure, since given the research context in which the therapy was delivered, standard clinical practices were not followed, and therefore it would not be valid to assess these phases of treatment on this sample. This selection process resulted in 20 recordings, 5 of which were used for the pilot trial and the remaining 15 being used to investigate interrater reliability.

Results

Interrater reliability is commonly measured using intraclass correlation coefficients (ICCs), which measure the degree to which scores within in the same data cluster or group resemble each other (Gwet, 2014). This differs from the percentage agreement method of reliability measurement in that the latter calculates the rate at which exact agreement between raters is achieved (Hallgren, 2012). Intraclass correlation is expressed as a coefficient (0–1) where higher scores reflect greater agreement. The consistency agreement will be used, which is the extent to which raters' scores can be expressed as linear functions of each other (as opposed to the extent to which scores absolutely agree).

Out of the 15 recordings assessed, 5 were incomplete in that the recordings had been terminated partway through the session. Table 1 shows descriptive statistics for each phase of the reprocessing fidelity checklist across both raters. The means indicate that fidelity was generally quite high for all recordings and the standard deviations indicate a general similarity within and between all rated scores. Except where noted, all data sets observed univariate normality by achieving nonsignificant Shapiro-Wilk's test statistics ($\alpha = 0.01$). For all 15 recordings, both raters assessed at least three of the six phases covered by the checklist. The phases most often omitted were Body Scan, Installation, and Closure. Body Scan and Installation require the complete processing of a target, so these phases were required far less frequently than the other phases. Closure was usually omitted from recordings due to lost recording data or time constraints from the therapist.

To estimate the interrater reliability of the reprocessing section of the EMDR Fidelity Checklist, two different types of intraclass correlation coefficients were computed. The first ICCs that we computed were based on the raters' assessments of fidelity *for each subscale*. For each subscale, we computed an ICC

TABLE 1. Subscale and Overall Descriptive Statistics for the A.5 EMDR Therapy Fidelity Checklist (Min = 0; Max = 2)

	Rater AS		Rater RC		Overall
Phase	*n*	Mean (*SD*)	*n*	Mean (*SD*)	Mean (*SD*)
III –Assessment	15	1.44 (0.46)	15	1.40 (0.48)	1.42 (0.46)
IV –Desensitization	15	1.18 (0.46)	15	1.21 (0.48)	1.20 (0.46)
V –Installation	5	1.45 (0.45)	4	1.57 (0.61)	1.50 (0.49)
VI –Body Scan	4	1.00 (0.82)	4	1.50 (1.00)	1.25 (0.89)
VII –Closure	9	1.31 (0.57)	10	1.51 (0.55)	1.41 (0.55)
VIII - Reevaluation	15	1.49 (0.55)	15	1.45 (0.35)	1.47 (0.46)
Overall	15	1.37 (0.42)	15	1.40 (0.40)	1.39 (0.40)

Note. *n* = number of sessions where this subscale was judged as applicable by rater (maximum 15); Mean = mean fidelity score; *SD* = standard deviation of fidelity scores; Overall scores comprised of fidelity checklist parameters drawn from all completed subscales.

based on the raters' relative agreement (i.e., agreement about the rank ordering across participants). Second, we computed, using the same procedure, the ICC for raters' assessment of the overall degree of fidelity, as assessed by the entire EMDR fidelity checklist (see Table 2). These ICCs used a two-way mixed-effects model with random people effects and fixed measures effects. To test the significance of these ICCs, a series of two-tailed *F* tests were conducted on these fidelity scores against a true value of zero (i.e., the null hypothesis ICC value) using an alpha level of .05. With the exception of the Body Scan subscale, which approached significance; $F[3, 3] = 9.00$, $p = .05$), all subscales exhibited significant ICCs (Reevaluation: $F[14, 14] = 4.29$, $p = .005$; Assessment: $F[14, 14] = 12.74$, $p < .001$; Desensitization: $F[14, 14] = 3.25$, $p = .02$; Installation: $F[3, 3] = 124.56$, $p = .001$; Closure: $F[8, 8] = 8.30$, $p = .004$). The ICC for the overall fidelity scores was also significant $F[14, 14] = 7.96$, $p < .001$. It should be noted here that the Body Scan subscale failed to observe univariate normality for one rater's score set.

Following this, the significant ICC values were ranked by strength to determine the degree to which interrater agreement was reached using this measure. Cicchetti (1994) generally ranks ICCs according to poor, fair, good, and excellent scores, with boundaries for these set at 0.40, 0.60, and 0.75, respectively. Here, it was found that all significant subscales obtained excellent ICCs, with the exception of the Desensitization subscale, whose ICC was within the good range (ICC = 0.69). Hence, an item-by-item analysis was conducted on the Desensitization subscale to improve its interrater reliability, which identified five low-performing items (i.e., items 16, 17, 19, 24a, and 24b; see Table 3).

TABLE 2. Interrater Agreement (ICC) for the Individual Subscale and Overall A.5 EMDR Therapy Fidelity Checklist

Phase	ICC	95% CI
III –Assessment	0.92***	[0.77, 0.97]
IV –Desensitization	0.81**	[0.43, 0.94]
V –Installation	0.99**	[0.88, 1.00]
VI –Body Scan	0.89	[-0.72, .99]
VII –Closure	0.88**	[0.47, 0.97]
VIII –Reevaluation	0.77**	[0.31, 0.92]
Total	.088***	[0.65, 0.96]

Note. *ICC* = intraclass correlation coefficient; *CI* = confidence interval; Total scores were calculated from fidelity checklist parameters scored across all completed subscales for each rater and for each recording. ICC for Desensitization subscale after improvements via item deletion and rescoring is shown above (original ICC = 0.69, 95% CI [0.08, 0.90], $F[14, 14] = 3.25$, $p = .02$). * $p < .05$. ** $p < .01$. *** $p < .001$.

By evaluating item-level ICCs for this subscale, items 16, 19, 24a, and 24b were identified as low-performing due to contributing zero variance to the subscale-level ICC and were subsequently eliminated from the final checklist (see Table 3). Item 17, which related to the amount of speech made by the administrator during the sets of bilateral movements (see Table 3), was also identified as low-performing (i.e., ICC = 0.064; $F[14, 14] = 1.07$, $p = .45$) but was instead recoded by the raters due to its relevance to EMDR protocol. Recoding was conducted by using more rigorous methods to more accurately determine the number of times EMDR administrators spoke. Independently recoding this item between the two raters greatly improved its interrater reliability (ICC = 0.93; $F[14, 14] = 13.46$, $p < .001$). Recalculating the

TABLE 3. Items Identified as Low-Performing After First ICC Analysis on Desensitization Subscale of Leeds' (2016) EMDR Fidelity Checklist (Appendix A.5)

Item	Question
16	Did the clinician provide bilateral eye movements or alternate bilateral stimulation of at least 24 to 30 repetitions per set as fast as could be tolerated comfortably? (Note: Children and adolescents and a few adult subjects require fewer passes per set, e.g., 14–20.)
17	During bilateral eye movements or alternate bilateral stimulation, did the clinician give some periodic nonspecific verbal support (perhaps contingent to nonverbal changes in subject) while avoiding dialogue?
19	After each verbal report, did the clinician promptly resume bilateral eye movements or alternate bilateral stimulation without excessive delay for discussion and without repeating subject's verbal report?
24a 24b	If subject showed extended intense emotion, did the clinician continue sets of bilateral eye movements or alternate bilateral stimulation with increased repetitions per set, remain calm and compassionate, and provide verbal cueing paced with the bilateral stimulation to encourage the subject to continue to "just notice" or "follow"? *(Skip if not applicable.* Counts as two items if applicable.)

Source: Leeds, A. M. (2016). *A guide to the standard EMDR protocols for clinicians, supervisors, and consultants* (2nd ed.). New York, NY: Springer Publishing.

Desensitization subscale using this modified data set resulted in a subscale with excellent interrater reliability, $F[14, 14] = 5.24$, $p = .002$; see Table 2. With these modifications, the reprocessing section of the EMDR Fidelity Checklist obtained an overall interrater reliability score which was well within the excellent range, $F[14, 14] = 8.39$, $p < .001$; see Table 2.

Missing Data

As mentioned above, five EMDR recordings were incomplete due to being terminated early. This primarily affected the evaluation of later items in the checklist, due to the chronological nature of the measure. Upon closer examination, the recordings all terminated during the Desensitization phase of the EMDR session, specifically making the recording inassessable after item 19 of the reprocessing checklist (i.e., "After each report, did the clinician promptly resume bilateral eye movements or alternate bilateral stimulation without excessive delay for discussion and without repeating the subject's verbal report?"). It is difficult to tell whether these incomplete recordings also created missing data for the Installation and Body Scan subscales as these are contingent upon the successful processing of a target, which occurs at the end of the Desensitization phase. Although it is clear that these incomplete recordings each created missing data for the Closure subscale, as this relates to a mandatory phase for the successful completion of a reprocessing session.

Discussion

The aim of this study was to pilot a comprehensive EMDR fidelity checklist for use by both researchers and clinicians and then assess its interrater reliability. The piloting stage involved a modified version of the reprocessing section of Leeds' (2016) EMDR Fidelity Checklist, which involved the creation of item-by-item scoring instructions, leading to a questionnaire that is easy to use by trained raters. With a few exceptions, the assessment stage yielded a set of ICCs which generally indicated excellent interrater reliability at both the subscale and overall level. This strongly indicates that scores on this questionnaire will likely be consistent between raters, which is a necessary criterion for use in both clinical and research contexts.

Although the average ICCs were in the excellent range, there was considerable variance in responses. At times, the lower level of the ICC even dipped into the poor-to-fair range (i.e., for the Desensitization, Body Scan, Closure, and Reevaluation subscales). Additional raters could help to strengthen the mean ratings; however, there are also some possibilities of improvements within each scale.

Due to the presence of incomplete recordings, the presence of a Closure phase often went unrecorded and therefore was not assessed by the raters, leading to fewer data-points and therefore more susceptible to score variation. This was not the case for the Reevaluation phase, which takes place at the beginning of a reprocessing session. Both scales could be assessed using more raters and on more clients to see if the mean increases or the variance in scores reduces.

A relatively simple interpretation exists for the nonsignificant ICC obtained by the Body Scan subscale. In this study, the Body Scan subscale was assessable for relatively few recordings in total. In addition, this subscale comprised the lowest number of items

(i.e., four), with only one mandatory item. Furthermore, this subscale marginally approached significance while obtaining a relatively high standard deviation. Hence, it is possible that the nonsignificant ICC achieved by the Body Scan subscale reflects the lack of statistical power that was available for calculating its score. This can be easily remedied in future studies by recalculating the ICC of this subscale with a higher number of data points, which may involve either the addition of more items to this subscale or otherwise scoring more EMDR recordings that contained the Body Scan phase.

The initial relatively lower ICC found for the Desensitization subscale seemed to occur with one item not performing as intended and four other items having little or no variance. Changing the response codes enabled one of the items to be retained and ICC improved. Although it remained necessary to remove the four items to achieve excellent interrater reliability, this is not to say that those items do not still have value. A problem was that therapists in the tapes assessed were both at an equivalent standard for each of these four items. If the checklist was used to assess a broader range of therapists, say, from at one end, people untrained and self-taught by reading an article on EMDR compared to expert therapists who were highly proficient, then these items may result in considerable variance and therefore ICC would increase. This could be tested in a future research project. Another argument to retain the current items is that for clinical purposes novice EMDR therapists might want to self-rate their sessions using this fidelity measure, and therefore paying attention to the particular aspect of EMDR as assessed by these items is important.

A methodical limitation of this study lies in that the raters that assessed the final scale also were involved in writing the item-by-item instructions for the reprocessing checklist. This potentially introduces a bias toward higher reliability scores in our study. The extent to which this was a limitation can be assessed in future research where interrater reliability scores can be evaluated using different raters.

The strength of this study is that to the authors' knowledge, this is the first time that an EMDR fidelity measure has been formally evaluated for reliability. The EFRS (Korn et al., 2017) also encompasses the reprocessing phases of EMDR with detailed ratings but has yet to be evaluated for interrater reliability.

The development of a comprehensive, strongly reliable, quantitative measure of EMDR fidelity represents an important preliminary step that opens up several possible avenues for future EMDR research. The scale should now be assessed in a larger sample to evaluate the degree that it predicts treatment outcome. Another future research issue is the degree of training required to provide ratings. In this study raters received a basic level (3-day training and had practiced the technique) but it may be possible, given the explicit nature of the checklist's items and ratings, that a lower level of training can still produce reliable results.

References

Adler-Tapia, R., & Settle, C. (2008). *EMDR and the art of psychotherapy with children*. New York, NY: Springer Publishing.

Adler-Tapia, R., & Settle, C. (2009). *EMDR in the treatment of childhood depression: Findings from a pilot study*. Unpublished manuscript.

Adler-Tapia, R., & Settle, C. (2016). *EMDR and the art of psychotherapy with children: Infants to adolescents* (2nd ed.). New York, NY: Springer Publishing.

Bassett, S. S., Stein, L. A., Rossi, J. S., & Martin, R. A. (2016). Evaluating measures of fidelity for substance abuse group treatment with incarcerated adolescents. *Journal of Substance Abuse Treatment, 66*, 9. doi:10.1016/j.jsat.2016.02.011

Borelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry, 71*(1), 52. doi:10.1111/j.1752-7325.2011.00233.x

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardised assessment instruments in psychology. *Psychological Assessment, 6*(4), 284. doi:10.1037/1040-3590.6.4.284

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. *American Journal of Community Psychology, 41*, 327. doi:10.1007/s10464-008-9165-0

Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review, 31*(1), 79. doi:10.1016/j.cpr.2010.09.007

Gwet, K. L. (2014). *Handbook of inter-rater reliability* (4th ed.). Gaithersburg, MD: Advanced. Analytics Press.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*, 23. doi:10.20982/tqmp.08.1.p023. Retrieved from http://www.tqmp.org/RegularArticles/vol08-1/p023/p023.pdf

Jeffries, F. W., & Davis, P. (2012). What is the role of eye movements desensitisation and reprocessing (EMDR) for post-traumatic stress disorder (PTSD)?: A review. *Behavioural and Cognitive Psychotherapy, 41*, 290. doi:10.1017/S1352465812000793

Korn, D. L., Maxfield, L., Smyth, N. J., & Stickgold, R. (2001). *EMDR fidelity rating scale (EFRS)*. Brookline, MA: Trauma Centre. Unpublished manuscript

Korn, D. L., Maxfield, L., Smyth, N. J., & Stickgold, R. (2017). *EMDR fidelity rating scale (EFRS)*. Retrieved from https://emdrresearchfoundation.org/emdr-fidelity-rating-scale

Laugharne, J., Kullack, C., Lee, C. W., McGuire, T., Brockman, S., Drummond, P. D., & Starkstein, M. D. (2016). Amygdala volumetric change following psychotherapy for posttraumatic stress disorder. *The Journal of Clinical Neuropsychiatry and Clinical Neurosciences, 28*(4), 312. doi:10.1176/appi.neuropsych.16010006

Lee, C., & Cuijpers, P. (2014). What does the data say about the importance of eye movement in EMDR? *Journal of Behavior Therapy and Experimental Psychiatry, 45*(1), 226. doi:10.1016/j.jbtep.2013.10.002

Leeds, A. M. (2016). *A guide to the standard EMDR protocols for clinicians, supervisors, and consultants* (2nd ed.). New York, NY: Springer Publishing.

Lu, W., Yanos, P. T., Gottlieb, J. D., Duva, S. M., Silverstein, S. M., Xie, H., . . . Mueser, K. T. (2012). Use of fidelity assessments to train clinicians in the CBT for PTSD program for clients with serious mental illness. *Psychiatric Services, 63*(8), 785. doi:10.1176/appi.ps.201000458

Maxfield, L., & Hyer, L. (2002). The relationship between efficacy and methodology in studies investigating EMDR treatment of PTSD. *Journal of Clinical Psychology, 58*, 23. doi:10.1002/jclp.1127

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75*(6), 829. doi:10.1037/0022-006X.75.6.829

Prowse, P. T., Meadows, G., & Enticott, J. (2015). An exploratory study into the effectiveness of fidelity scales in the delivery of mindfulness-based cognitive therapy. *Mindfulness, 6*(6), 1401. doi:10.1007/s12671-015-0412-9

Schoenwald, S. K. (2011). It's a bird, it's a plan, it's . . . fidelity measurement in the real world!. *Clinical Psychology: Science and Practice, 18*(2), 142. doi:10.1111/j.1468-2850.2011.01245.x

Shapiro, F. (2001). *Eye movement desensitization and reprocessing: Basic principles, protocols and procedures* (2nd ed.). New York, NY: Guilford Press.

Shapiro, F. (2018). *Eye movement desensitization and reprocessing: Basic principles, protocols and procedures* (3rd ed.). New York, NY: Guilford Press.

Southam-Gerow, M. A., McLeod, B. D., Arnold, C. C., Rodriguez, A., Cox, J. R., Reise, S. P., . . . Kendall, P. C. (2016). Initial development of a treatment adherence measure for cognitive-behavioral therapy for child anxiety. *Psychological Assessment, 28*(1), 70. doi:10.1037/pas0000141

Van der Kolk, B., Spinazzola, J., Blaustein, M., Hopper, J., Hopper, E., Korn, D., & Simpson, W. (2007). A randomised clinical trial of EMDR, fluoxetine and pill placebo in the treatment of PTSD: Treatment effects and long-term maintenance. *Journal of Clinical Psychiatry, 68*, 37. doi:10.4088/JCP.v68n0105

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behavioural Research & Therapy, 47*(2), 119. doi:10.1016/j.brat.2008.10.018

Disclosure. In the article we recommend that a degree of standard EMDR training is required to understand the items in the scale, although we remain open to the possibility that this is not the case. Two of the authors (AL and CL) provide basic EMDR trainings and therefore may benefit financially from providing such training sessions. The authors wish to declare this as a conflict of interest in publishing this study.

Correspondence regarding this article should be directed to Christopher W. Lee, School of Psychiatry, University of Western Australia, 35 Stirling Highway, Crawley, Perth WA, Australia 6009. E-mail: chris.lee@uwa.edu.au

Appendix. Final Fidelity Checklist

EMDR Therapy Fidelity Rating Scale for Reprocessing Session			
Subject Code		Date of Session:	
Rater:		Date of Review:	
Comments:		Average Rating:	

Reevaluation Phase average score (items 1–4):	
Assessment Phase average score (items 5–14):	
Desensitization Phase average score (items 15–28):	
Installation Phase average score (items 29–34):	
Body Scan Phase average score (items 35–38):	
Closure Phase average score (items 39–45):	

	Reevaluation Phase			
1	Did the clinician reevaluate the subject's experience since the last session with attention to feedback from the log, presenting complaints, responses to current stimuli, and additional memories or issues that might warrant modifications to the treatment plan? (This is crucial after history-taking sessions as well as after stabilization and reprocessing sessions.) 0—Clinician never or minimally elicited subject's progress on these areas. 1—Clinician elicited subject's progress on these areas in an incomplete or fundamentally flawed manner (e.g., spending an hour on this activity, eliciting lots of irrelevant information, failing to fully explore relevant issues). 2—Clinician elicited subject's progress on these areas well.	**0**	**1**	**2**
2	Did the clinician check the SUD and VoC on the target from the last session? (*Skip if this is the first reprocessing session.*) 0—Clinician checks neither SUD nor VoC. 1—Clinician checks either SUD or VoC. 2—Clinician checks both SUD and VoC.	**0**	**1**	**2**
3	Did the clinician check for additional aspects of the target from the last session that may need further reprocessing? (*Skip if this is the first reprocessing session.*) *Examples include "When you think of that image, what's the worst part of it now?" or "Has that image or any related thoughts or feelings been bothering you since we last met?"* 0—Clinician never explored this. 1—Clinician explored this in an incomplete or fundamentally flawed manner (e.g., asked "Have you been getting any flashbacks?") 2—Clinician explored this well.	**0**	**1**	**2**

(*continued*)

(Continued)

	Reevaluation Phase			
4	If the target from the last session had been incomplete or if in this session the subject reported the SUD were now a 1 or above or the VoC were a 5 or below, did the clinician resume reprocessing on the target from the last session? (*Skip if this is the first reprocessing session. If the client has multiple traumas and after reprocessing the SUDS is a 2 or even a 3, it may be more appropriate to target a more disturbing or related memory or earlier memory, then select this as the next target.)* 0—Reprocessing was evidently incomplete but the clinician did not remain focused on this target (i.e., chose a new target, ended the session). 1—Reprocessing was evidently incomplete but clinician chose to focus on an associated memory. 2—Reprocessing was evidently incomplete and clinician chose to remain focused on this target.	**0**	**1**	**2**
	Reevaluation Phase average score (items 1–4): Possible total of four items. Three items (2, 3, and 4) can be skipped before reprocessing sessions have begun.			

	Assessment Phase			
5	Did the clinician select an appropriate target from the treatment plan? 0—No target was selected. 1—Selected target was irrelevant to presenting problems and case formulation OR was fundamentally flawed in some way (e.g., was not a sensory event). 2—Selected target was relevant and appropriate.	**0**	**1**	**2**
6	Did the clinician elicit a picture (or other sensory memory) that represented the entire incident or the worst part of the incident? 0—Clinician did not elicit a sensory representation of the event. 1—Clinician elicited a sensory representation of the event in a fundamentally flawed way (e.g., selected multiple representations at once, chose the most tolerable sensory representation). 2—Clinician elicited and chose an appropriate sensory representation of the event.	**0**	**1**	**2**
7	Did the clinician elicit an appropriate negative cognition (NC)? 0—NC is not obtained or is suggested by clinician and does not appear to resonate with subject. 1—NC is missing a couple of essential elements. 2—NC is derived from the subject and is self-referencing, presently held, accurately focuses on presenting issue, generalizable, is a true cognition (i.e., not a feeling, like "I am frustrated"), and has affective resonance.	**0**	**1**	**2**
8	Did the clinician elicit an appropriate positive cognition (PC)? 0—PC is not obtained or is suggested by clinician and does not appear to resonate with subject. 1—PC is missing a couple of essential elements. 2—PC is derived from the subject and is self-referencing, in the same theme as the NC, accurately focuses on desired direction of change, generalizable, is a true cognition (i.e., not a feeling, like "I am happy"), is realistically adaptive, and 1 < VoC < 5.	**0**	**1**	**2**

(*continued*)

(Continued)

	Assessment Phase			
9	Did the clinician assure that the NC and PC address the same thematic domain: responsibility, safety, choice? 0—NC and PC are in different thematic domains. 1—NC and PC did not clearly address the same thematic domain. 2—NC and PC clearly addressed the same thematic domain.	**0**	**1**	**2**
10	Did the clinician obtain a valid VoC by referencing the felt confidence of the PC in the present while the subject focused on the picture (or other sensory memory)? 0—VoC is absent or invalid (i.e., VoC < 1 or VoC > 5). 1—Valid VoC obtained but not while focused on image or other sensory memory OR invalid VoC obtained while focusing on image or other sensory memory. 2—Valid VoC obtained while focusing on image or other sensory memory.	**0**	**1**	**2**
11	Did the clinician elicit the present emotion by linking the picture and the NC? 0—Did not elicit the present emotion (or physiological response). 1—Elicited present emotion (or physiological response) from the image or the NC but not both. 2—Elicited present emotion (or physiological response) from both the image and the NC.	**0**	**1**	**2**
12	Did the clinician obtain a valid SUD (i.e., the current level of disturbance for the entire experience—not merely for a present emotion)? NB: SUD rating is on the entire target experience. 0—Did not obtain a SUD. 1—SUD obtained but not valid (i.e., SUD <= 2 during a first processing session, although continuing with a SUD <= 2 may be appropriate during a reprocessing session). 2—Valid SUD obtained on present emotion (or physiological response).	**0**	**1**	**2**
13	Did the clinician elicit a body location for current felt disturbance? 0—Did not elicit a body location for current disturbance. 1—Elicited a vague body location for current disturbance. 2—Elicited body location for current disturbance.	**0**	**1**	**2**
14	Did the clinician follow the standard assessment sequence listed above? *Note: Although some leeway on the standard sequence is acceptable during this phase, the sequence of eliciting the Image → NC → PC → VoC →* Emotion → SUD → Location *is essential because the subject may find it difficult to elicit a PC after eliciting the current emotion associated with the traumatic event.* 0—Did not follow the essential sequence of Image → NC → PC → VoC → Emotion → SUD → Location. 1—Mostly followed the essential sequence of Image → NC → PC → VoC → Emotion → SUD → Location. 2—Followed the essential sequence of Image → NC → PC → VoC → Emotion → SUD → Location.	**0**	**1**	**2**
Assessment Phase average score (items 5–14): Total of 10 items.				

Desensitization Phase				
15	Before beginning bilateral eye movements or alternate bilateral stimulation, did the clinician instruct subject to focus on the picture, NC (in the first person), and the body location? 0—Did not instruct subject to focus on any of these areas. 1—Clinician instructed subject to focus on one or two items (image or sensory memory, NC, and body location). 2—Clinician instructed subject to focus on all three items (image or sensory memory, NC, and body location).	**0**	**1**	**2**
16	Did the clinician provide bilateral eye movements or alternate bilateral stimulation of at least 24 to 30 repetitions per set as fast as could be tolerated comfortably? (Note: Children and adolescents and a few adult subjects require fewer passes per set, e.g., 14–20.) 0—Did not administer any bilateral eye movements or alternate bilateral stimulation (EM/ABS) or offered a speed of stimulation that was significantly too slow or far too few repetitions (e.g., only four to eight saccades). 1—Most times, most sets missing an essential element of EM/ABS, somewhat too slow or somewhat too few saccades. 2—Most times, most sets were at least 24 EM/ABS of relatively constant and sufficient speed, width, and direction.	**0**	**1**	**2**
17	During bilateral eye movements or alternate bilateral stimulation, did the clinician give some periodic nonspecific verbal support (perhaps contingent to nonverbal changes in subject) while avoiding dialogue? 0—Gave no nonspecific verbal support or was overly directly with specific feedback or excessive dialogue during most sets (i.e., spoke during >50% of the set). 1—Gave limited nonspecific verbal support or only slightly overly specific feedback or excessive dialogue during some of the sets (i.e., <50% of the set). 2—Most time, most sets avoided excessive dialogue and specific feedback and did offer nonspecific verbal support (i.e., if subject is not emotional, at least one comment per set; if subject is emotional, then more frequently).	**0**	**1**	**2**
18	At the end of each discrete set of bilateral eye movements or alternate bilateral stimulation, did the clinician use appropriate phrases to have the subject, "Rest, take a deeper breath, let it go"(while not asking the subject to "relax") then make a *general* inquiry ("What do you notice now?") while avoiding narrowly *specific* inquiries about the image, emotions, or feelings? 0—Used inappropriate phrases after most sets (i.e., >50% of the set). 1—Used inappropriate phrases after some sets (i.e., <50% of the set). 2—The clinician used appropriate phrases for all three items after most sets, most of the time (i.e., deep breath instruction, general inquiry, avoided specific inquiry).	**0**	**1**	**2**
19	After each verbal report, did the clinician promptly resume bilateral eye movements or alternate bilateral stimulation without excessive delay for discussion and without repeating subject's verbal report? 0—Permitted or encouraged excessing verbal reports or needlessly repeated subject's comments after some sets (i.e., >50% of the sets). 1—Often resumed EM/ABS without repeating the subject's verbal report and without promoting excess verbiage (i.e., <50% of the sets). 2—Completed the above most of the time, after most sets.	**0**	**1**	**2**

(*continued*)

(Continued)

	Desensitization Phase			
20	If verbal reports and nonverbal observations indicated reprocessing was effective, after reaching a neutral or positive channel end, did clinician return attention to the selected target and check for additional material in need of reprocessing (i.e., "What's the worst part of it now?")? 0—Subject was never asked a question similar to "Recall the original incident. What do you notice now?" after reaching a neutral or positive end without evidence of strengthening. 1—After five or more consecutive sets of EM/ABS reporting neutral or positive experiences without evidence of strengthening, only then was the subject asked a question similar to "Recall the original incident. What do you notice now?" 2—After two consecutive sets of EM/ABS reporting neutral or positive experiences without evidence of strengthening, subject was asked a question similar to "Recall the original incident. What do you notice now?"	**0**	**1**	**2**
21	If verbal reports or nonverbal observations indicated reprocessing was ineffective, did the clinician vary characteristics of the bilateral eye movements or alternate bilateral stimulation (speed, direction, change modality, etc.)? (*Skip if not applicable*. Counts as two items if applicable.) 0—After three or four consecutive sets of eye movements reporting no change in a memory, belief, emotion, or body location, clinician never made a valid variation of the EM/ABS. 1—After three or four consecutive sets of eye movements reporting no change in a memory, belief, emotion, or body location, clinician made a valid variation of the EM/ABS. 2—After two consecutive sets of eye movements reporting no change in a memory, belief, emotion, or body location, clinician made a valid variation of the EM/ABS.	**0**	**1**	**2**
22	If verbal reports or nonverbal observations indicated reprocessing were ineffective, did the clinician do any of the following? *(Skip if not applicable.* Counts as two items if applicable.) 1. Explore for an earlier disturbing memory with similar affect, body sensations, behavioral responses, urges, or belief. 2. Explore for a blocking belief, fear, or concern disrupting effective reprocessing, and then identify a related memory. 3. Explore target memory for more disturbing images, sounds, smells, thoughts, beliefs, emotions, or body sensation. 4. Invite subject to imagine expressing unspoken words or acting on unacted urges. 5. Offer one or more interweaves. 0—After two consecutive sets of eye movements reporting no change in a memory, belief, emotion, or body location, clinician did not try any of these strategies. 1—After two consecutive sets of eye movements reporting no change in a memory, belief, emotion, or body location, clinician didn't persist in using one of the above strategies (i.e., tried one strategy but subject still blocked, and didn't try a second strategy). 2—After two consecutive sets of eye movements reporting no change in a memory, belief, emotion, or body location, clinician effectively used one or more of these strategies.	**0**	**1**	**2**

(*continued*)

(Continued)

	Desensitization Phase			
23	If subject showed extended intense emotion, or if reprocessing was ineffective, did clinician show appropriate judgment in selecting and offering one (or if necessary more) interweave(s) from among the categories of responsibility, safety, and choices while avoiding excess verbiage? *(Skip if not applicable.* Counts as two items if applicable.) *Note: Intense, extended emotion includes a single behavior (e.g., crying, hyperventilating, trembling, turning red, or other more subtle signs as determined by the therapist) that is present for an extended time (i.e., >6 minutes). Ineffective processing is when the subject reports exactly the same experience (e.g., emotion, thought, image, or body disturbance) OR a repetitive set of responses (i.e., looping) after two or more successive sets.* 0—Clinician did not use an interweave where appropriate. 1—Interweave was offered in an incomplete or fundamentally flawed manner (e.g., interweave took 10 minutes to deliver, interweave was not from domains of responsibility, safety, or choice). 2—An interweave from the domains of responsibility, safety, or choice was offered in an appropriate way.	**0**	**1**	**2**
24	If subject showed extended intense emotion, did the clinician continue sets of bilateral eye movements or alternate bilateral stimulation with increased repetitions per set, remain calm and compassionate, and provide verbal cueing paced with the bilateral stimulation to encourage the subject to continue to "just notice" or "follow"? *(Skip if not applicable.* Counts as two items if applicable.) *Note: Intense, extended emotion includes a single behavior (e.g., crying, hyperventilating, trembling, turning red) that is present for an extended time (i.e., >6 minutes).* 0—Clinician did not increase repetitions per set or give calm, compassionate, and encouraging verbal cueing. 1—Clinician either increased repetitions per set until emotional behavior noticeably decreased OR gave limited calm, compassionate, and encouraging verbal cueing (but not both). 2—Clinician increased repetitions per set until emotional behavior noticeably decreased AND gave multiple calm, compassionate, and encouraging verbal cueing per set.	**0**	**1**	**2**
25	If a more recent memory emerged, did the clinician acknowledge its significance, offer to return to the more recent memory later, and redirect the client back to the selected target memory within one or two sets of bilateral eye movements or alternate bilateral stimulation? *(Skip if not applicable.)* 0—A recent memory emerged and clinician did not acknowledge its significance or offer to return to it later, but merely continued with many sets (more than four or five) of EM/ABS focused on the recent memory without returning to check the original target memory. A significant portion of the remaining portion of the session continued with this new focus of attention. 1—A recent memory emerged and clinician either acknowledged its significance while offering to return to it later OR redirected subject's attention to target memory (but not both) within two or three sets of EM/ABSs. Alternatively, recent memory emerged and clinician both acknowledged its significance while offering to return to it later AND redirected subject's attention to target memory, but did so after more than three but fewer than 6 sets of EM/ABS. 2—Recent memory emerged and all components of this item (i.e., acknowledgment, redirection to target, responding within two EM/ABS) were achieved completely.	**0**	**1**	**2**

(*continued*)

(Continued)

	Desensitization Phase			
26	If an earlier (antecedent) memory emerged, did the clinician continue bilateral eye movements or alternate bilateral stimulation on the earlier memory, and if this earlier memory became resolved, did the clinician redirect the subject back to the target memory? Alternatively, did the clinician make a clinically informed decision to help the subject to contain this material until a later date due to concerns that the subject was not ready to confront this material? *(Skip if not applicable.)* If earlier memory *did not* require immediate containment: 0—Clinician did not offer EM/ABS until earlier memory was resolved. Instead the clinician immediately redirected the subject to the original target even though time remained to process the earlier memory. 1—Clinician offered EM/ABS for a series of sets after which the subject reported neutral or positive experiences, but they never redirected subject's attention back to the original target. 2—Clinician offered EM/ABS until the subject reported neutral or positive experiences and if time remained then redirected the subject's attention to back to the original target. If earlier memory *did* require prompt containment (this may not be evident immediately): 0—Clinician never advised the subject to about the option to contain this material and did not explore with the subject whether to address this earlier material now or wait until a later date when they feel more ready to confront it. 1—Clinician delayed their advice to the subject to contain this material until a later date and the subject subsequently requested to stop reprocessing after confronting the earlier memory. Alternatively, clinician promptly advised the subject to contain this material without giving the subject the option of continuing, or may not have stated when they would return to it or the reasons for doing so. 2—Clinician explored with the subject the option to contain this material until a later date when they are able to confront it and the subject elected to contain it.	**0**	**1**	**2**
27	If it became clear it was not possible to complete reprocessing in this session, did clinician show appropriate judgment to avoid returning subject's attention to residual disturbance in target, skip Installation and Body Scan phases, and go directly to closure? (*Skip if not applicable.*) *Note: Clinicians should make this decision within 10 minutes of the session ending. This decision is informed partly by clinical judgment and partly by the subject's reported SUD upon rechecking the target after two sets of their reporting positive or neutral experiences. The aim is to ensure that subjects are oriented to the present and are given enough time to regain full orientation to the present, and to diminish any residual anxiety and distress before leaving the session.* Reprocessing evidently could not be completed in this session and: 0—The clinician never made any decision in order to end the session effectively and continued reprocessing right up to the end of the session. 1—The clinician made some decisions in order to end the session effectively, however these were delayed, incomplete, rushed, or otherwise fundamentally flawed. (e.g., beginning part of the installation phase first and then going directly to closure; not reserving sufficient time for closure based on the client's needs). 2—The clinician went directly to Closure phase without returning the subject's attention to the residual disturbance in target.	**0**	**1**	**2**

(*continued*)

(Continued)

	Desensitization Phase			
28	If it appeared from spontaneous subject reports that the Desensitization Phase may have been complete, did clinician show appropriate judgment to return subject's attention to target to confirm the SUD was 0 (or an "ecological" 1) by offering at least one more set of bilateral eye movements or alternate bilateral stimulation on the target before going to the Installation Phase? (*Skip if not applicable.) Target was checked (e.g., by asking, "Recall the original incident. What do you notice now?") AND:* 0—Appropriate SUD was not obtained before moving onto Installation Phase. 1—Appropriate SUD was obtained but not rechecked after a second set of EM/ABS before moving onto Installation Phase. 2—Appropriate SUD was obtained and rechecked after (at least) a second set of EM/ABS before moving onto Installation Phase.	**0**	**1**	**2**
Desensitization Phase average score (items 15–28): Up to eight items can be skipped. Fourteen items, plus four can be doubled.				

	Installation Phase			
If the Desensitization Phase was completed (and item 28 was scored) proceed to score Installation Phase items. If the Desensitization Phase was incomplete, skip both the Installation and Body Scan Phases and proceed to score the Closure Phase. However, if the desensitization was incomplete and the clinician incorrectly proceeded to the Installation or Body Scan Phases, these phases should be scored and down rated accordingly.				
29	Did the clinician confirm the final PC by inquiring whether the original PC still fit or if there were now a more suitable one? 0—Clinician did not check to see if a better PC could be elicited and merely began Installation with the original PC from Phase 3. 1—Clinician inquired about the a better PC but began the Installation Phase with a final PC that did not match full criteria for a PC or that was not a good fit for the subject. 2—Clinician checked to see if a better PC could be elicited began the Installation Phase with a final PC that the subject agreed was suitable and that fully matched criteria for a PC.	**0**	**1**	**2**
30	Before offering bilateral eye movements or alternate bilateral stimulation, did the clinician obtain a valid VoC (i.e., by having subject assess the felt confidence of the PC while thinking of the target incident)? 0—Subject was never prompted for a VoC. 1—Subject was not instructed to think about the target incident before providing a VoC for the PC. Alternately, EM/ABS began before subject gave a valid VoC. 2—Subject was instructed to think about target incident before providing a VoC for the PC (and before being administered the EM/ABS).	**0**	**1**	**2**
31	Did the clinician offer more sets of bilateral eye movements or alternate bilateral stimulation after first asking each time that the subject focus on the target incident and the final PC? 0—Subject was not given a series of EM/ABS or alternately, subject was never instructed to focus on both the target incident and the PC between each set of EM/ABS. 1—Subject was instructed to focus on either the target incident or the PC (but not both) between sets EM/ABS. 2—Subject was instructed to focus on both target incident and PC between sets of EM/ABS.	**0**	**1**	**2**

(*continued*)

(Continued)

	Installation Phase			
32	Did the clinician obtain a valid VoC after each set of bilateral eye movements or alternate bilateral stimulation? 0—Clinician failed to obtain a valid VoC after more than half of all EM/ABS sets. 1—Clinician obtained a valid VoC after more than half but not all EM/ABS sets. 2—Clinician obtained a valid VoC after all EM/ABS sets.	**0**	**1**	**2**
33	After sets of bilateral eye movements or alternate bilateral stimulation, if the VoC did not rise to a 7, did the clinician inquire what prevents it from rising to a 7 and then make an appropriate decision to target the thought or move to body scan or closure? (*Skip if not applicable.) VoC was struggling to rise to a 7 after several sets of eye movements and:* 0—Clinician did not make the inquiry as per above. 1—Clinician made an inquiry and accepted the subject's rationale for the VoC remaining below a 7 without targeting the rational with further EM/ABS. 2—Clinician made the inquiry as per above and appropriately targeted the thought or moved to Body Scan / Closure.	**0**	**1**	**2**
34	Did the clinician continue sets of bilateral eye movements or alternate bilateral stimulation until the VoC was a 7 and no longer getting stronger (or a 6 if "ecological")? *(Skip if not applicable.) (Note either item 33 or 34 should be scored unless there were [a]insufficient time to complete the Installation Phase or [b]a new issue emerged that prevented completing the Installation Phase.)* 0—The completion of the Installation Phase did not involve the use of VoCs. 1—The completion of the Installation Phase involved the incomplete or fundamentally flawed use of VoC's (e.g., ending with a single VoC of 7, ending with two successive VoC's of 5). 2—The completion of the Installation Phase occurred via obtaining VoCs of 7 (or "ecological" 6's) after two successive sets of EM/ABS.	**0**	**1**	**2**
Installation Phase average score (items 29–34): Up to two items can be skipped. Possible total six items.				

	Body Scan Phase			
35	Did the clinician obtain a valid body scan (asking subject to [a] report any unpleasant sensation while focusing on [b] the final PC and [c] the target incident with eyes closed)? 0—No body scan was conducted. Or the subject was asked to think about negative details from the sensory memory, emotions or physical sensations in Phase 3. 1—A body scan was conducted, but subject was not instructed to focus on *both* the final PC and the target incident. 2—Subject was instructed on all major components of body scan.	**0**	**1**	**2**
36	*If any unpleasant sensations were reported,* did the clinician continue with additional sets of bilateral eye movements or alternate bilateral stimulation until these sensations became neutral or positive? If unpleasant sensations were reported and bilateral stimulation was not offered, was there an appropriate clinical rationale (i.e., linkage to a different memory)? *(Skip if not applicable.) Unpleasant sensations were reported and:* 0—No additional sets of EM/ABS were offered and no appropriate clinical rationale was present. 1—Additional sets of EM/ABS were offered and were discontinued before the subject reported neutral or positive experiences after two successive sets. 2—Additional sets of EM/ABS were offered and were discontinued after the subject reported neutral or positive experiences after two successive sets. Alternatively, No additional sets of EM/ABSs were offered but an appropriate clinical rationale was present.	**0**	**1**	**2**

(*continued*)

(Continued)

	Body Scan Phase			
37	*If a new memory emerged*, did the clinician make an appropriate decision to continue by targeting the new memory in the session or later as part of the treatment plan? *(Skip if not applicable.) Note: The new memory must be an eligible target (i.e., it must relate to presenting problems and have some distressing content). A new memory emerged and:* 0—The clinician neither targeted it in session (i.e., starting from Phase 3) nor explained to the subject that it may be best to target it later in treatment. 1—The clinician either targeted it in session (i.e., starting from Phase 3) or explained to the subject that it may be best to target it later in treatment, however the decision made was not well-informed by the session's remaining time or the nature of the memory. 2—The clinician either targeted it in session (i.e., starting from Phase 3) or explained to the subject that it may be best to target it later in treatment. This decision was well-informed by the session's remaining time and the nature of the memory.	**0**	**1**	**2**
38	*If pleasant sensations were reported*, did the clinician target these and continue with additional sets of bilateral eye movements or alternate bilateral stimulation as long as these sensations continued to become more positive? *(Skip if not applicable.)*	**0**	**1**	**2**
	Body Scan Phase average score (items 35–38): Up to three items can be skipped. Possible total of four items.			

	Closure Phase			
39	Did the clinician make an appropriate decision to move to closure? 0—The Closure Phase was omitted. 1—The Closure Phase began prematurely or was delayed. 2—The Closure Phase was begun in a timely manner from either the successful completion of the Body Scan Phase or an appropriate premature discontinue from an earlier phase due to time or distress management constraints.	**0**	**1**	**2**
40	Did the clinician assure subject was appropriately reoriented to the present by (a) *assessing* subject's residual distress and to enhance orientation to the present and (b) *if needed* then offer appropriate and sufficient structured procedures (such as guided imagery, breathing exercises, or containment exercise to decrease anxiety, distress, and dissociation, 0—Subject was not assessed for distress and clinician continued immersive discussion of the memory. When needed, interventions were not used to diminish the subject's distress. 1—Subject was assessed for distress, but attempts at orienting them to the present and diminishing their distress were incomplete or ineffective. 2—Subject was assessed for distress and clinician began present-oriented discussion. When needed, interventions were used to diminish subject's distress and subject reported these to be effective.	**0**	**1**	**2**
41	Did the clinician support mentalization by inviting subject to comment on changes in awareness, perspective, and self-acceptance related to the session just completed? 0—No discussion about the subject's in-session experiences, the treatment trajectory, or observed improvements occurred. 1—Some comments about the session's in session experiences, the treatment trajectory, or observed improvements occurred. 2—Considered discussion about the subject's in-session experiences, the treatment trajectory, or observed improvements occurred.	**0**	**1**	**2**

(*continued*)

(Continued)

	Closure Phase			
42	Did the clinician offer empathy and psychoeducation where appropriate, and statements to normalize and help to put into perspective the subject's experience? *(Skip if not applicable.)* 0—Subject introduced information about their own experiences, the treatment trajectory, and/or presenting problems and clinician did not respond therapeutically. 1—Subject introduced information about their own experiences, the treatment trajectory, and presenting problems and clinician gave partially therapeutic responses. 2—Subject introduced information about their own experiences, the treatment trajectory, and presenting problems and clinician responded with empathy, normalizing statements, or psychoeducation.	**0**	**1**	**2**
43	Did the clinician brief the subject on the possibility between sessions of continuing or new, positive, or distressing thoughts, feelings, images, sensations, urges, or other memories or dreams related to the reprocessing from this session? 0—Clinician did not brief the subject of this possibility. 1—Clinician minimally briefed the subject of this possibility. 2—Clinician fully (and concisely) briefed the subject of this possibility.	**0**	**1**	**2**
44	Did the clinician request that the subject keep a written log of any continuing or new issues or other changes to share at the next session? 0—Clinician did not request that subject keep written notes of any between-session behavioral observations, insights, triggers, etc. 1—Clinician requested that subject keep notes of between-session issues or observations in an incomplete or fundamentally flawed manner, i.e. without explaining the notes can be brief and/or without offering a written log form. 2—Clinician requested that subject keep notes of between-session issues in a complete manner, e.g. explaining that they could be about behavioral changes, responses to triggers, new insights, new memories, positive dreams, or nightmares.	**0**	**1**	**2**
45	Did the clinician remind the subject to practice a self-control procedure daily or as needed? 0—Clinician did not remind the subject to practice self-control procedures. 1—Clinician reminded subject to practice self-control procedures in an incomplete or fundamentally flawed manner. 2—Clinician reminded subject to practice self-control procedures.	**0**	**1**	**2**
Closure Phase average score (items 39–45): Total of seven items. One item #42 may be skipped.				