



A Novel Approach to the Assessment of Fidelity to a Cognitive Behavioral Therapy for PTSD Using Clinical Worksheets: A Proof of Concept With Cognitive Processing Therapy

Shannon Wiltsey Stirman

Dissemination and Training Division, National Center for PTSD, VA Palo Alto Healthcare System & Stanford University

Cassidy A. Gutner

Women's Health Sciences Division, National Center for PTSD, VA Boston Healthcare System & Boston University School of Medicine

Jennifer Gamarra

University of California, Los Angeles

Michael K. Suvak

Suffolk University

Dawne Vogt

Women's Health Sciences Division, National Center for PTSD, VA Boston Healthcare System & Boston University School of Medicine

Clara Johnson

Dissemination and Training Division, National Center for PTSD

Jennifer Schuster Wachen

Women's Health Sciences Division, National Center for PTSD, VA Boston Healthcare System & Boston University School of Medicine

Katherine A. Dondanville

University of Texas Health Science Center at San Antonio

Jeffrey S. Yarvis

Carl R. Darnall Army Medical Center, Fort Hood, TX

Jim Mintz

University of Texas Health Science Center at San Antonio

Alan L. Peterson

University of Texas Health Science Center at San Antonio
South Texas Veterans Health Care System
University of Texas at San Antonio

Stacey Young-McCaughan

University of Texas Health Science Center at San Antonio

Patricia A. Resick

Duke University Medical Center

for the STRONG STAR Consortium

Fidelity monitoring is a critical indicator of psychotherapy quality and is central to successful implementation. A major barrier to fidelity in routine care is the lack of feasible, scalable, and valid measurement strategies. A reliable, low-burden fidelity assessment would promote sustained implementation of cognitive behavioral therapies (CBTs). The current study examined fidelity measurement for cognitive processing therapy (CPT) for posttraumatic stress disorder (PTSD) using clinical worksheets. External raters evaluated patient worksheets done as a part of treatment, both guided by the therapist and completed independently as homework. Results demonstrated that fidelity ratings from CPT session worksheets were feasible and efficient. Notably, they were strongly correlated with observer ratings of the fidelity of CPT strategies that were present on the worksheets. Agreement among ratings conducted by individuals with a range of experience with CPT was acceptable to high. There was not a main effect of therapist-guided, in-session worksheet ratings on PTSD symptom change. However, patient competence in completing worksheets independently was associated with greater PTSD symptom decline and in-session, therapist-guided worksheet completion was associated with larger symptom decreases among patients with high levels of competence. With further research and refinement, rating of worksheets

Funding for this work was made possible by the U.S. Department of Defense through the U.S. Army Medical Research and Materiel Command, Congressionally Directed Medical Research Programs, Psychological Health and Traumatic Brain Injury Research Program awards W81XWH-08-02-109 (Alan Peterson), W81XWH-08-02-0114 (Brett Litz) and W81XWH-08-02-0116 (Patricia Resick). Additional support was received from National Institute of Mental Health award #R21 MH112099169 (Shannon Wiltsey Stirman).

The grant sponsor played no role in study design; the collection, analysis, and interpretation of data; the writing of this paper; or the decision to submit this paper for publication.

The views expressed herein are solely those of the authors and do not reflect an endorsement by or the official policy or position of the U.S. Army, the Department of Defense, the Department of Veterans Affairs, or the U.S. Government.

This paper is dedicated to the memory of Michael K. Suvak, who passed away on June 4, 2020.

Address correspondence to Shannon Wiltsey-Stirman, Ph.D., National Center for PTSD, VA Palo Alto Healthcare System, 795 Willow Road, Menlo Park, CA 94025; e-mail: Shannon.Wiltsey-Stirman@va.gov

may be an efficient way to examine therapist and patient skill in key CPT elements, and their interactions, compared to the gold standard of observer ratings of therapy video-recordings. Additional research is needed to determine if worksheets are an accurate and scalable alternative to gold standard observer ratings in settings in which time and resources are limited.

Keywords: Fidelity; Cognitive Behavioral Therapy; Cognitive Processing Therapy; PTSD

POLICYMAKERS HAVE INCREASINGLY EMPHASIZED the importance of assessing both the use and the quality of evidence-based psychosocial and mental health interventions. Both the Mental Health Parity & Addiction Equity Act of 2008 and the Patient Protection and Affordable Care Act of 2010 highlight the role of evidence-based treatment strategies in service delivery and call for attention to quality assessment and accountability. Additionally, fidelity must be assessed for research and practical efforts to implement effective mental health interventions (Proctor et al., 2011). Fidelity includes adherence to an intervention protocol, competence or skill of delivery, treatment differentiation, and treatment receipt, or the patients' ability to understand and use the intervention (Dusenbury et al., 2003; Schoenwald et al., 2011), and it is considered a key indicator of quality of psychotherapeutic interventions (Fairburn & Cooper, 2011). Research indicates that fidelity is associated with greater symptom improvement in cognitive therapies, including cognitive processing therapy (CPT) for posttraumatic stress disorder (PTSD; Farmer et al., 2017; Marques et al., 2019) and cognitive therapy for depression (Feeley et al., 1999; Strunk et al., 2010).

Given its importance, ongoing fidelity monitoring and support is a key component in evidence-based psychotherapy implementation and sustainability models (Aarons et al., 2011; Feldstein et al., 2008; Shediach-Rizkallah & Bone, 1998). The possibility of "voltage drops," or reductions in effectiveness, is of particular concern in these efforts, as it is important to determine whether decreases in

treatment effectiveness are due to aspects of treatment delivery such as treatment fidelity, as opposed to differences in setting or participant characteristics (Chambers, 2012; Schoenwald et al., 2011). As large mental health systems have implemented evidence-based psychotherapies (Clark, 2011; Karlin & Cross, 2014; Nakamura et al., 2011), the need to develop scalable fidelity assessment has increased.

Many efforts to assess fidelity to evidence-based psychotherapies during initial training efforts have employed instruments developed for the assessment of fidelity in clinical trials. These efforts nearly always involve direct observation of sessions or review recorded sessions (Creed et al., 2014; Karlin et al., 2012). While considered the gold-standard in psychotherapy outcome research, strategies that require review of session recordings are time and labor intensive and not feasible in many practice settings (Schoenwald et al., 2011; Wiltsey Stirman et al., 2015), particularly in the context of long-term efforts to support quality delivery of evidence-based psychotherapies. Sufficient personnel and funding for session observation are unlikely to be allocated in large systems with thousands of therapists to monitor. They also are unlikely to be available in smaller, underresourced settings.

Other approaches for monitoring fidelity have involved either therapist or patient reports and/or review of clinical progress notes. However, research has identified limitations to these indirect methods of monitoring psychotherapy fidelity. Some studies have found that therapist self-reports of fidelity only modestly correspond with observer or patient fidelity ratings (Carroll et al., 2010; Decker et al., 2013). Moreover, because therapists often perceive a self-report of their fidelity to treatment as adding to an already high paperwork burden, they may over- or underreport their use of prescribed treatment elements or not complete the reports as directed (Kauth et al., 2010). While therapists may be able to report somewhat accurately on their own adherence to treatment, they may have difficulty reporting on their own skill level (Brosan et al., 2008), particularly with regard to more nuanced aspects of the treatment such as behavioral rehearsal of cognitive skills (Ward et al., 2013). Clinical progress notes have also been found to have limited utility for assessing fidelity. Research indicates that progress notes provide poor estimates of frequency or intensity of evidence-based treatment techniques (Jensen-Doss et al., 2008; Liddle et al., 2006). Even if progress notes, which document a therapist's perspective of what occurred in sessions, could provide a sufficiently valid indication of adherence, they are unlikely to

accurately reflect competence levels for session elements. Likewise, research on patient reports of therapist's fidelity suggests that in addition to increasing patient burden, patients may not fully understand therapy elements or notice changes in adherence, and measures may be subject to ceiling effects (Schoenwald et al., 2011). Patient recall may also be influenced by timing of administration or the complexity of the behaviors being assessed (Hrisos et al., 2009).

One source of information from cognitive behavioral therapy (CBT) that could shed light on both adherence and competence is the CBT worksheet. CBT worksheets that are completed or reviewed in session contain details about aspects of cognitive restructuring that might otherwise only be obtained from listening to an entire session; thus, they may be leveraged to provide a more efficient way to evaluate fidelity. In many forms of CBT, therapists use worksheets to guide the patient in identifying and challenging maladaptive beliefs, two important treatment strategies that can lead to symptom reduction for many disorders (e.g., Clark et al., 2014; Resick et al., 2008; Vogt et al., 2012). Because worksheets map closely to these cognitive elements, worksheets completed in session with therapist guidance can provide evidence of adherence. The quality of the thoughts, feelings, and elements of cognitive restructuring entered on worksheets completed in sessions can also provide evidence of therapist competence, or the quality with which these elements of CBTs were addressed within the session. For example, common errors that are apparent on worksheets may include incorrectly identifying automatic thoughts and/or including other negative automatic thoughts in the cognitive restructuring or rational response sections. The therapist's lack of attention to errors on "homework" worksheets that are reviewed in session may indicate that the therapist is not adequately monitoring and addressing their patients' ability to use CBT strategies independently (i.e., treatment receipt; Borrelli et al., 2005; Johnson-Kozlow et al., 2008). Skilled completion of homework may be necessary to promote clinical change (Wiltsey Stirman, Gutner, et al., 2018). Thus, compared to other fidelity-monitoring approaches, clinical worksheets completed both in session and as homework may provide a less labor-intensive, but rich and potentially accurate, source of information about the fidelity with which certain cognitive elements of a session were implemented within session.

The primary purpose of this study was to examine whether clinical worksheets could be used to assess therapist fidelity to treatment by

examining the feasibility and acceptability of this method, interrater reliability, and criterion-related validity. We hypothesized that a system for coding worksheets would be feasible (e.g., worksheets would be legible and complete enough to rate; time to rate worksheets would be less than the time to observe and rate full sessions; and the difficulty of using the system would be low). Analyses of criterion-related validity examined the association between worksheet fidelity and observer-based ratings of recorded sessions, with the expectation that there would be a high association among these measures. We also sought to explore the relationship between the competence ratings and subsequent symptom change as the existing literature is mixed (Webb et al., 2010). We hypothesized that fidelity ratings based on therapist-guided, in-session worksheets (hereafter described as in-session worksheets) would be associated with subsequent symptom change. We also hypothesized that an interaction between in-session worksheet quality and patient skill at completing worksheets independently would be associated with subsequent symptom change.

Method

DEVELOPMENT OF THE RATING SYSTEM

Subject matter experts with knowledge of both CPT and the systems in which it is implemented, along with experts in measurement, assisted in the development and refinement of a scoring

system and manual (available from first author). Worksheets from a previous clinical trial testing CPT (Resick et al., 2002) were used in an iterative process to refine a scoring system and decision rules to score each element (varying between 3 and 8 elements per worksheet) of each worksheet for adherence (a 0-1 rating, indicating whether or not it had been completed), and competence (a 0-2 rating, with a 2 indicating fully correct or skilled completion). The study team and subject matter experts then scored worksheets that were collected from a different trial testing CPT (Resick et al., 2008) prior to the commencement of rating for the current study. An overall adherence and competence rating for each session was computed by taking the mean of scores for each worksheet element (3-8 per worksheet) on all worksheets rated for that session. A mean score for one of each type of worksheet that was completed for homework after each session was also computed.

INTERVENTION AND MATERIALS

After the development and training of the scoring system, raters coded worksheets collected from individual therapy sessions conducted as part of a randomized clinical trial (RCT) testing CPT in a military setting (Resick et al., 2017). CPT is a 12-session, trauma-focused cognitive therapy that teaches patients to examine and change problematic beliefs about their role in the trauma, themselves, and the world that were altered as

Table 1
Clinical Materials Collected to Assess Fidelity in Sessions of Cognitive Processing Therapy

Session	Material (when completed)	Purpose
2	Impact statement (HW) Stuck Point Log (IS) ABC (IS)	HW-assess whether and to what extent the patient was able to respond to prompts in the assignment regarding reasons for the trauma and its impact IS-assess therapist ability to identify stuck points during impact statement processing IS-assess therapist ability to help patient differentiate events, thoughts and feelings
3	ABC (HW and IS)	HW-assess patient ability to differentiate events, thoughts and feelings IS-assess therapist ability to help patient differentiate events, thoughts and feelings
4	ABC (HW) Challenging Questions (IS)	HW-assess patient ability to differentiate events, thoughts and feelings IS-assess therapist ability to help patient challenge assimilated stuck points
5	Challenging Questions (HW and IS) Patterns of Problematic Thinking (IS)	HW-assess patient ability to challenge assimilated stuck points IS-assess therapist ability to help patient challenge assimilated stuck points IS-assess therapist ability to help patient identify thought patterns
6	Patterns of Problematic Thinking (HW and IS) Challenging Beliefs (IS)	HW-assess patient ability to identify thought patterns IS-assess therapist ability to help patient identify thought patterns IS-assess therapist ability to help patient challenge stuck points and identify alternative beliefs
7-12	Challenging Beliefs (HW and IS)	HW-assess patient ability to challenge stuck points and identify alternative beliefs IS-assess therapist ability to help patient challenge stuck points and identify alternative beliefs

Note. HW = homework completed by patient; IS = completed in session with therapist guidance; ABC = ABC Worksheet.

a result of traumatic events (Resick et al., 2008). The efficacy of CPT has been established in group (Resick & Schnicke, 1992; Resick et al., 2017) and individual formats (Resick et al., 2008; Resick et al., 2002; Resick et al., 2017), and in a variety of traumatized populations (Bass et al., 2013; Galovski et al., 2013; Monson et al., 2006; Resick et al., 2008). Like other CBTs, CPT uses progressive clinical worksheets at every session to illustrate the concepts being presented and provide patients with opportunities to practice cognitive skills. The worksheets are introduced at every session and completed together by the therapist and patient in session. Worksheets are then assigned as homework for the patient to complete independently and are reviewed and corrected by the therapist in subsequent sessions. Table 1 describes the materials that are collected during therapy and their purposes.

PARTICIPANTS

The worksheets were obtained from 106 active-duty military members who received individual CPT and completed worksheets as part of a larger trial examining group vs. individual CPT (Resick et al., 2017). All participants had a diagnosis of PTSD according to the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR; American Psychiatric Association, 2000)*. See Table 2 for descriptive statistics from the 93 participants with homework and in-session worksheet ratings from

at least one session between Sessions 4 and 7 who were included in the current analyses examining worksheet competence scores as predictors of PTSD symptom change.

ASSESSMENT INSTRUMENTS

PTSD Assessment

Patients participated in the Posttraumatic Symptom Scale–Interview Version (PSS-I) prior to randomization. The PSS-I is a 17-item clinical interview that evaluates *DSM-IV* PTSD symptoms on a frequency and severity scale (scores range, 0–51, with higher scores reflecting greater PTSD severity). The stressor-specific Posttraumatic Stress Disorder Checklist (PCL-S) is a self-reported measure of PTSD symptoms in the past month (scores range, 17–85, with higher scores reflecting greater PTSD severity). While the baseline measure assessed symptoms over the past month, the version administered during treatment queried symptoms over the previous week (Weathers et al., 1993).

Fidelity Ratings

Worksheet rating. When completing and collecting the worksheets in the study, therapists indicated whether the worksheet was completed in session by the therapist and patient, completed for homework and reviewed in session, or completed by the patient for homework but not reviewed in session. To rate fidelity to treatment within session, one of each of the worksheets that was completed or reviewed in session was scored to be representative of that session. Raters chose the worksheet that was most complete, or if all were similarly complete, one that was most closely related to the goals and theme of that particular CPT session. Patient skill as a proxy for therapist fidelity was assessed by rating the skill with which they completed worksheets independently between sessions, using only worksheets that were not reviewed by the therapist. To assess each type of worksheet and corresponding skills, homework assignments from Sessions 1–11 were rated for each patient. No worksheets are completed in Session 1 of the protocol, but patients are given an assignment that was collected in Session 2 for rating, which was used to reflect the patients' level of skill after the first session.

Two expert CPT providers (both rostered by the CPT for PTSD program; and one of whom was also a qualified CPT consultant), two bachelor's-level research assistants, and one undergraduate student served as raters. Individuals of different clinical backgrounds were included as raters to explore the feasibility of rating fidelity based on worksheets. To ensure all raters understood CPT and its core elements, three nontherapists read the CPT manual and completed a Web-based training on CPT. The

Table 2
Patient Demographics Included in Current Study

Variable	(<i>n</i> = 93) No. (%) of Patients
Age, mean (<i>SD</i>), y	33.5 (7.2)
Time in service, mean (<i>SD</i>), y	11.5 (6.9)
No. of deployments, mean (<i>SD</i>)	2.3 (1.1)
Gender	
Male	91.4%
Female	8.6%
Race	
White	41.9%
Black or African American	35.5%
Other	22.6%
Ethnicity	
Hispanic or Latino	22.5%
Education	
High school diploma/ GED	16.1%
Some college or more	82.8%
Hazardous drinking	15.1%
History of head injury with postconcussive symptoms	71.8%
Combat-related index trauma	88.2%

Note. This sample consists of patients with homework and in-session worksheet ratings from at least one session between sessions 4 and 7. No. = number; *SD* = standard deviation; y = years.

group of five raters reviewed the coding system and manual together before any rating began. The raters independently rated worksheets from two cases per week over 4 weeks and met to discuss discrepancies in ratings, clarifying decision rules in the scoring manual throughout this process. When raters achieved 90% agreement, they began coding independently. All raters rated an even number of worksheets with some overlapping to assess agreement. Fifteen percent of the 1,582 sessions evaluated were coded by two raters to assess for agreement. To prevent drift, raters met every other week to discuss any discrepancies in ratings on overlapping cases during the rating period. After scoring worksheets for each session, raters reported worksheet legibility, time required for ratings, how accurately the worksheets seemed to reflect patient and therapist skill, how accurately they seemed to reflect what occurred during the session, and the ease of completion of the ratings.

Observer ratings. All sessions in the CPT study were videotaped for supervision and traditional fidelity ratings. CPT therapists who were not involved in the current project rated CPT adherence and competence in past trials and in the parent study as observer raters of fidelity. The raters followed a version of the CPT Fidelity Rating form that has been used in past studies (Peterson et al., 2013; Resick et al., 2017). This involved the use of dichotomous variables for adherence to each unique and essential item within sessions as well as a 5-point Likert-type scale for competence at each unique and essential treatment element. As described in greater detail in the supplement report from the parent study (Resick et al., 2017) raters assessed fidelity from one or more individual therapy sessions from 103 different patients, with a total of 107 different session ratings. Unique and essential elements were rated as present in 97.5% of the ratings. The mean of the scores for each unique and essential item for each session was calculated to determine an overall score for adherence and competence for each session. In the overall trial, the average individual therapist competence score was 4.4 out of 5.0 ($SD = 0.7$). We also calculated a separate session-level mean that included only observer rating items that corresponded to activities captured on the worksheets (e.g., introducing a particular worksheet and guiding the patient through it, and completing cognitive restructuring activities). Items such as psychoeducation or assigning homework were not included in this analysis. For the current study, a subsample of 44 session videos that were reviewed aligned with worksheets that were coded.

DATA ANALYSES

Feasibility and Acceptability

Frequencies, percentages, and descriptive statistics were calculated to determine perceived feasibility, efficiency, and acceptability of the scoring strategy.

Reliability and Criterion Validity

Rater agreement from the subsample of worksheets scored twice was assessed for individual worksheet scores and for overall session scores using Cohen's kappa for adherence ratings, while intraclass correlation coefficients (ICCs) were calculated for competence ratings using two way mixed models. We selected absolute agreement for a more stringent model of rater reliability. We examined correlations between fidelity ratings assessed using worksheets and the observer-rated fidelity assessments from the original clinical trial.

Exploration of Relationship Between Fidelity Ratings and Symptom Change

With the assumption that high fidelity treatment will result in reductions in patient reports of PTSD symptoms regardless of other factors, we examined the impact of competence on change in PTSD symptoms from assessment to assessment using latent change score (LCS) modeling and the Mplus software package (Version 7; Muthén & Muthén, 1998-2012). The LCS approach simultaneously models two change processes (See Figure 1). Autoregressive relationships (also referred to as regressed-change models) are evaluated by regressing each assessment of a repeatedly measured outcome variable on the previous assessment of the same variable (McArdle, 2009). LCS models also include a general growth factor to account for changes in mean levels and variability in the outcome across time (i.e., LCS models do not assume equal variance of the repeatedly assessed variable over time), thereby producing more reliable and accurate estimates of change between assessments. In addition, the LCS approach disaggregates within and between subjects' variability in both the mean level of and change in the outcome variable, reducing the likelihood of biasing that can occur due to the confounding of within- and between-subjects effects (Conklin & Strunk, 2015; Hamaker et al., 2015).

To explore associations between in-session and homework competence ratings and PTSD symptoms over time, we entered competence scores as predictors of each latent change variable from the session at which these variables were assessed through the follow-up assessment. Competence scores were examined in three separate models, one each for competence at Session 2, at Session 3, and between Sessions 4 and 7 (see results section for rationale for

these groupings). Because each latent change score has paths from subsequent PTSD levels and the general growth factor, each latent change score isolates change during that specific period (i.e., change when controlling for previous levels and general change processes). We evaluated two models for each set of competence score predictors. The first model included in-session and patient competence scores to evaluate main effects. The second model added an interaction (product) term to evaluate interactions between the two scores.

Results

FEASIBILITY, ACCEPTABILITY, AND RELATIVE ADVANTAGE
 Of 621 sessions that could potentially be rated, 83% could be rated for adherence and competence.

Fewer than 10% had legibility issues. This suggests that capturing data generated during CPT is a feasible strategy. Raters indicated that the strategy was low to moderately difficult to use, with a mean rating of 2.6 ($SD = 1.6$) on a 0-6 scale (with 0 indicating “no difficulty” and 6 indicating “not possible to rate”) of 1,582 worksheets. They believed that the rating system was moderately effective at capturing session content ($M = 3.2$, $SD = 1.4$) of 570 worksheets and that their ratings were moderately reflective of therapist skill ($M = 3.3$, $SD = 1.4$) of 574 worksheets and patient skill ($M = 3.3$, $SD = 1.1$) of 572 worksheets. The mean time to complete ratings for a session (one or more worksheet per session) was 7.0 minutes, ($SD = 3.9$) for 1,582 worksheets.

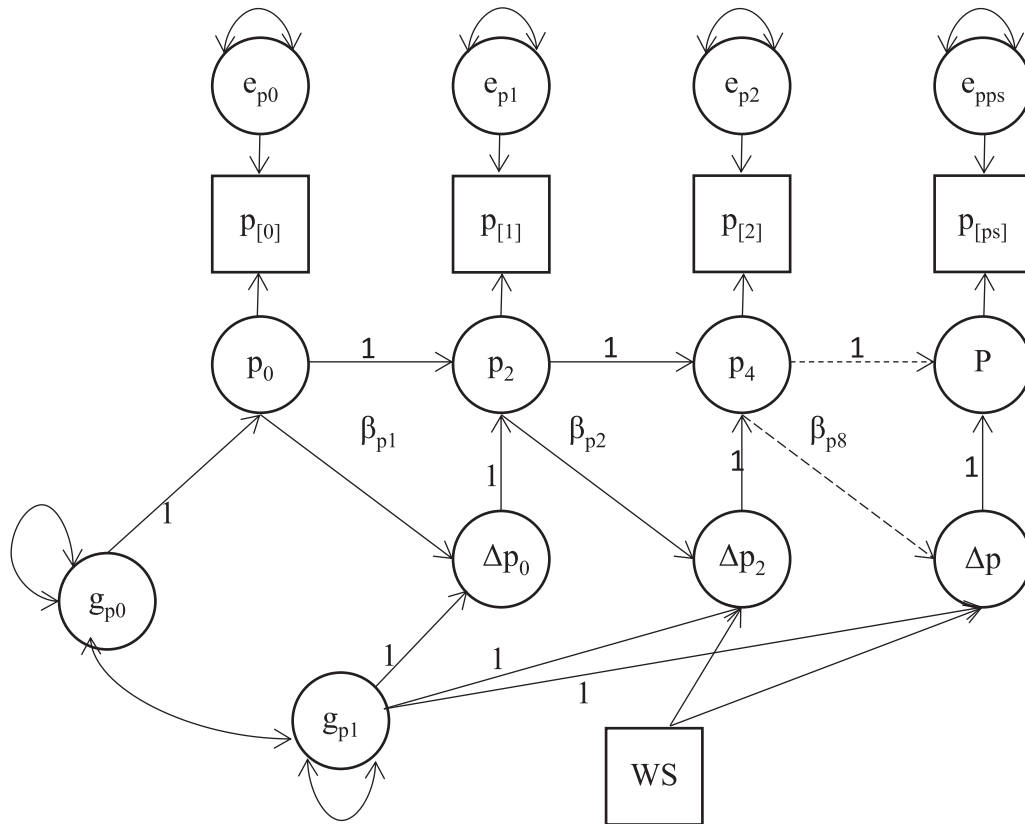


FIGURE 1 Depiction of dual change LCS model for PTSD symptoms (p) symptoms. Boxes represent measured variables (e.g., the box with $p[0]$ represents the score of PCL-S at the pretreatment assessment). Circles represent latent variables with p_0 (latent variable of PTSD at baseline), Δp_{02} (latent change score for PTSD between pre-treatment and session 2), g_{p0} (intercept or latent variable representing pre-treatment value), g_{p1} (constant change factor), and e_p (measurement residual/error) used as the notation to depict the different types of latent variables in the model. Its represent the proportional change coefficients (i.e., the influence of previous level on subsequent change). Double-sided arrows indicated nondirection (i.e., correlational) associations, while single-sided arrows depict directional relationships. Eight assessments (pretreatment, Sessions 2, 4, 6, 8, 10, 12, and posttreatment) were used for the analyses. However, to save space, only four assessments are depicted in the figure; therefore, the dashed lines indicate that the figure is discontinuous, going from the fourth assessment to the eighth assessment. LCS = latent change score; PTSD = posttraumatic stress disorder; PCL-S = PTSD Checklist – Stressor-Specific.

Table 3
Descriptive Statistics for Worksheet Collected and Rater Agreement Statistics

Worksheet	In-Session Adherence				In-Session Competence				Patient Out-of-Session Homework Competence			
	<i>n</i>	<i>M</i> (<i>SD</i>)	Min- Max	Rater Agreement (<i>k</i>)	<i>n</i>	<i>M</i> (<i>SD</i>)	Min- Max	Rater Agreement (ICC)	<i>n</i>	<i>M</i> (<i>SD</i>)	Min- Max	Rater Agreement (ICC)
ABC Sheet	217	.98 (.10)	0-1	.98	198	1.73 (.25)	1-2	.68	180	1.58 (.42)	0-2	.55
Impact Statement/Stuck Point Log	109	.97 (.14)	0-1	.60	85	1.78 (.36)	0-2	.89	93	1.45 (.60)	0-2	.68
Challenging Questions	135	1 (0)	1-1	.96	125	1.54 (.39)	.5-2	.76	101	1.32 (.47)	0-2	.88
Problematic Patterns	90	.96 (.15)	0-1	.99	77	1.55 (.54)	0-2	.63	74	1.39 (.62)	0-2	.92
Challenging Beliefs	144	.95 (.09)	.56-1	.76	132	1.43 (.33)	.4-2	.79	133	1.27 (.44)	0-2	.76
Overall Session Adherence/Competence	599	.97 (.09)	0-1	.93	516	1.60 (.36)	0-2	.86	493	1.40 (.49)	0-2	.88

Note. *k* = Cohen's Kappa; ICC = intraclass correlation coefficient.

RELIABILITY AND VALIDITY

Interrater Reliability

Table 3 contains descriptive statistics for each worksheet rating and rater agreement statistics. Reliability ratings between the five study raters were in the good to almost perfect range for rating worksheets evaluating therapist adherence and competence, and agreement on worksheets completed in session adherence was excellent (Cicchetti, 1994). Ratings of patient skill completing both in-session and homework worksheets ranged from adequate to almost perfect, with excellent agreement for the session's overall skill rating.

Internal Consistency Reliability and Correlations Between Sessions

Table 4 depicts the zero-order correlations among baseline PTSD scores and in-session and homework worksheet competence ratings of Sessions 2 through 7. In-session and homework worksheet competence were correlated in Sessions 2-6. In-session worksheet competence at Session 2 (ABC worksheets, which identify thoughts and feelings), only exhibited a significant negative association with in-session worksheet competence at Session 7 (Challenging Beliefs worksheets), while in-session worksheet competence at Session 3 (ABC worksheets) was not associated with in-session competence at any other session. In contrast, in-session worksheet competence scores between Sessions 4 through 7 (worksheets that challenge beliefs) tended to be significantly and positively associated with each other. A similar pattern emerged for homework worksheet competence ratings; therefore, to reduce the number

of analyses, we analyzed worksheet competence ratings from Session 2 and Session 3 separately, while taking the average of worksheet competence ratings from Sessions 4-7. Cronbach's α were .68 and .60 respectively for in-session and homework worksheet competence ratings from Sessions 4-7, indicating acceptable levels of internal reliability for a 4-item scale, given that alphas are typically lower for scales with fewer items.

Criterion-Related Validity

Correlation between expert observers' fidelity ratings of videotaped treatment sessions ($n = 44$ sessions) and adherence as rated by coding worksheets was low for adherence when the full observer score was compared to the rated adherence of worksheets ($r = .06, p = .68$) but was higher with the when the observer score was compared to the scores of worksheet-related items within the sessions ($r = .81, p < .001$). The correlation between fidelity ratings of session videos and session competence from worksheet ratings was not significant for full sessions ($r = .28, p = .07$), but worksheet-related items were highly correlated with observer ratings of items that were reflected on worksheets ($r = .82, p < .001$). These findings suggest that the worksheet rating strategy may not be appropriate to assess fidelity of a full session, but may adequately reflect fidelity to cognitive elements of the session involving essential skills and elements that are captured on worksheets.

ASSOCIATIONS BETWEEN WORKSHEETS AND SYMPTOM CHANGE

PTSD Univariate LCS Model

Prior to examining associations between worksheet competence and change in PTSD, we evaluated

Table 4
Descriptive Statistics and Correlations

Variable	1	2	3	4	5	6	7	8	9	10	11	12	<i>M</i>	<i>SD</i>	
1. PCL PRE	—												55.03	11.00	
2. S_WS_Com2	.18	—											1.77	0.26	
3. S_WS_Com3	-.23*	.10	—										1.72	0.26	
4. S_WS_Com4	-.16	.12	.20	—									1.59	0.34	
5. S_WS_Com5	.04	.02	.09	.28*	—								1.55	0.42	
6. S_WS_Com6	.01	.00	.02	.37**	.34**	—							1.52	0.39	
7. S_WS_Com7	.02	-.29*	-.07	.14	.33**	.32**	—						1.39	0.35	
8. PT_WS_Com2	.01	.35**	.15	.15	.04	-.08	-.06	—					1.50	0.49	
9. PT_WS_Com3	-.10	.14	.25*	.26*	-.07	-.03	-.07	.25*	—				1.66	0.28	
10. PT_WS_Com4	-.10	.19	.20	.38**	.33**	.13	-.08	.27*	.32**	—			1.40	0.47	
11. PT_WS_Com5	-.14	.38**	.12	.31**	.5**	.46**	.13	.21	.02	.30*	—		1.34	0.55	
12. PT_WS_Com6	-.13	.03	.09	.26*	.29*	.36**	.35**	-.01	.15	.35**	.21	—	1.33	0.43	
13. PT_WS_Com7	.02	.03	.07	.19	-.05	.12	.22	.23	.17	.35**	.13	.51**	—	1.20	0.48

Note. Pearson's correlations presented. *M* = mean, *SD* = standard deviation; PCL = Posttraumatic Checklist; PRE = pre-treatment; S = Session (e.g., in-session, therapist guided); WS = worksheet; PT = patient; Com = competency (with adjacent number = session number). * $p < .05$. ** $p < .01$.

a univariate LCS model for PTSD symptoms. This model included proportional change estimates (i.e., paths from PTSD levels at Time = t to latent change variables representing change in PTSD from Time = t to Time = $t + 1$), a general growth factor, and a specified autocorrelated error structure among the residuals of the repeated PTSD measurements. The fit indicators showed adequate to good model fit, $\chi^2(30, N = 93) = 54.25, p = .004$; *CFI* = .96, *TLI* = .96; *RMSEA* = .09, *SRMR* = .06. Next, the proportional change coefficients were constrained to be equal across time to test whether the impact of PTSD symptoms at one point on subsequent change in PTSD between that point and the next time point was invariant across time. Table 5 presents the estimates of change between each PTSD assessment and corresponding 95% confidence intervals. On average, participants exhibited significant decreases between each PTSD assessment except between Session 12 and the posttreatment assessment. The estimate for the intercept indicated that on average participants' baseline PTSD (PCL) score was 54.03 ($cr = 47.63, p < .001$), and the variance estimate for the intercept term was 42.59, ($cr = 2.43, p = .02$), indicating significant individual differences in initial status. The estimate for the general growth factor was 16.00 ($cr = 3.46, p = .001$) with the variance estimate for the growth factor indicating significant variance across participants in general growth ($b = 28.80, cr = 2.43, p = .02$).

Associations Between Worksheet Scores and Symptom Change

Table 6 summarizes the results of the three LCS models that included competence scores as predic-

tors of PTSD latent change scores. Although we expected that in-session competence would be associated with subsequent symptom change, findings did not support this hypothesis. A significant main effect of patient competence completing homework worksheets emerged, with higher patient competence at Session 2 significantly associated with changes in PTSD from Session 8 to Session 10 ($b = -4.99, cr = -2.89, p = .004$). Specifically, patients who were rated higher in competence on their independently completed Session 2 homework (identifying events, thoughts and feelings) exhibited larger decreases in PTSD from Session 8 to Session 10 (challenging beliefs; $d = -.63$) than patients whose Session 2 homework was rated lower in competence. For Session 3 competence scores, results indicated a main effect of patient competence on PTSD changes from Session 8 to Session 10 ($b = 7.51, cr =$

Table 5
PTSD Change Scores from Assessment to Assessment

Time	Δ PTSD (95%CI)
Period	
PRE-S2	-2.52 (-4.54, -0.50)
S2-S4	-1.79 (-3.43, -0.14)
S4-S6	-2.11 (-3.65, -0.57)
S6-S8	-1.91 (-3.48, -0.34)
S8-S10	-2.45 (-4.11, -0.79)
S10-S12	-2.58 (-4.37, -0.80)
S12-PST	-0.51 (-2.38, 1.35)

Note. Δ PTSD = change in PTSD; 95%CI = 95% confidence interval; PRE = pretreatment assessment; S = session; PST = posttreatment assessment.

Table 6
Worksheet Fidelity Predicting Changes in PTSD

Path	b	CR	p	d	Path	b	CR	p	d	Path	b	CR	p	d
<u>WSF(2) → ΔPTSD 2-4</u>														
Main Eff.					WSF(3) → ΔPTSD 2-4					WSF(4-7) → ΔPTSD 2-4				
Session	0.25	0.08	.937	0.02	Main Eff.	-3.12	-0.92	.357	-0.20					
Patient	2.19	1.04	.297	0.23	Interactions	0.35	0.11	.910	0.02					
<u>WSF(2) → ΔPTSD 4-6</u>														
Main Eff.					WSF(3) → ΔPTSD 4-6					WSF(4-7) → ΔPTSD 4-6				
Session	-0.43	-0.14	.889	-0.03	Main Eff.	-4.39	-1.30	.195	-0.28					
Patient	0.26	0.16	.872	0.04	Interactions	-0.11	-0.03	.976	-0.01					
Interactions					Interactions									
Session	-0.27	-0.07	.941	-0.02	Session	-5.00	-1.63	.102	-0.36					
Patient	0.20	0.12	.906	0.03	Patient	-1.25	-0.32	.748	-0.07					
Interactions					Interaction	-14.65	-1.04	.297	-0.23					
<u>WSF(2) → ΔPTSD 6-8</u>														
Main Eff.					WSF(3) → ΔPTSD 6-8					WSF(4-7) → ΔPTSD 6-8				
Session	3.72	1.32	.187	0.29	Main Eff.	-1.12	-0.35	.726	-0.08	Main Eff.	1.06	0.33	.745	0.07
Patient	0.07	0.05	.963	0.01	Session	0.09	0.03	.978	0.01	Session	-2.87	-1.41	.158	-0.29
Interactions					Patient					Patient				
Session	3.62	1.10	.269	0.24	Interactions					Interactions				
Patient	-0.17	-0.12	.903	-0.03	Session	-1.52	-0.47	.637	-0.10	Session	3.11	0.83	.408	0.17
					Patient	-0.50	-0.152	0.879	-0.03	Patient	-3.77	-1.85	.065	-0.38

(continued on next page)

Path	b	CR	p	d	Path	b	CR	p	d	Path	b	CR	p	d
Interaction	-1.43	-0.28	.779	-0.06	Interaction	-2.69	-0.223	0.823	-0.05	Interaction	11.39	1.87	.062	0.39
<u>WSF(2) → ΔPTSD 8-10</u>					<u>WSF(3) → ΔPTSD 8-10</u>					<u>WSF(4-7) → ΔPTSD 8-10</u>				
Main Eff.					Main Eff.					Main Eff.				
Session	0.12	0.03	.975	0.01	Session	-3.33	-0.89	.373	-0.19	Session	-1.36	-0.35	.727	-0.07
Patient	-4.99	-2.89	.004	-0.63	Patient	7.51	2.18	.029	0.48	Patient	-1.87	-0.57	.572	-0.12
Interactions					Interactions					Interactions				
Session	-3.41	-0.93	.352	-0.20	Session	-3.50	-0.92	.356	-0.20	Session	-5.80	-1.53	.126	-0.32
Patient	-4.15	-2.62	.009	-0.57	Patient	7.39	2.266	0.023	0.49	Patient	-1.41	-0.44	.660	-0.09
Interaction	-20.83	-3.56	.000	-0.78	Interaction	5.54	0.422	0.673	0.09	Interaction	22.55	-2.86	.004	-0.59
<u>WSF(2) → ΔPTSD 10-12</u>					<u>WSF(3) → ΔPTSD 10-12</u>					<u>WSF(4-7) → ΔPTSD 10-12</u>				
Main Eff.					Main Eff.					Main Eff.				
Session	4.12	1.25	.211	0.27	Session	-1.51	-0.36	.721	-0.08	Session	5.24	1.59	.112	0.33
Patient	-0.19	-0.09	.928	-0.02	Patient	0.92	0.26	.797	0.06	Patient	-5.07	-1.53	.127	-0.32
Interactions					Interactions					Interactions				
Session	6.02	1.64	.102	0.36	Session	-2.27	-0.532	0.595	-0.12	Session	4.99	1.09	.276	0.23
Patient	-1.24	-0.61	.543	-0.13	Patient	-0.40	-0.122	0.903	-0.03	Patient	-5.82	-1.76	.079	-0.36
Interaction	8.30	1.36	.175	0.30	Interaction	-12.53	-0.993	0.321	-0.22	Interaction	-0.12	-0.01	.991	0.00
<u>WSF(2) → ΔPTSD 12-PST</u>					<u>WSF(3) → ΔPTSD 12-PST</u>					<u>WSF(4-7) → ΔPTSD 12-PST</u>				
Main Eff.					Main Eff.					Main Eff.				
Session	2.86	0.726	.468	0.16	Session	1.20	0.27	.789	0.06	Session	3.36	1.00	.318	0.21
Patient	0.44	0.214	.831	0.05	Patient	-3.08	-0.68	.495	-0.15	Patient	2.06	0.68	.495	0.14
Interactions					Interactions					Interactions				
Session	0.66	0.17	.864	0.04	Session	-0.03	-0.01	.995	0.00	Session	1.99	0.52	.607	0.11
Patient	0.33	0.20	.846	0.04	Patient	-5.15	-1.11	.268	-0.24	Patient	1.23	0.38	.704	0.08
Interaction	-15.26	-3.06	.002	-0.67	Interaction	-19.01	-1.46	.143	-0.32	Interaction	-6.00	-1.07	.285	-0.22

Note. b = unstandardized regression coefficient; CR = critical ratio (estimate/standard error); p = p value; d = Cohen's d where $d = 2 * CR / n$; WSF = worksheet fidelity; Main Eff. = main effect; ΔPTSD = change in PTSD; PST = posttreatment assessment.

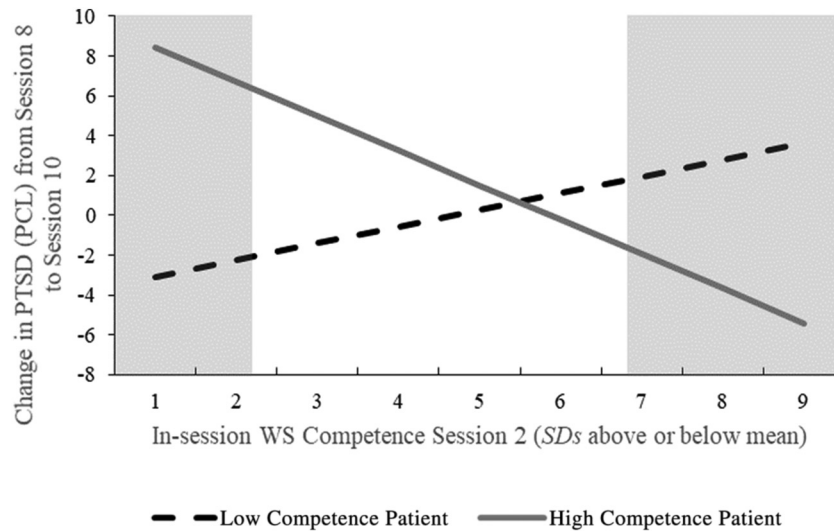


FIGURE 2 The impact of in-session, therapist-guided worksheet competence (Session 2) on change in PTSD from Session 8 to Session 10 as a function of homework worksheet competency (Session 2). PTSD = Posttraumatic Stress Disorder; PCL = PTSD Checklist – Stressor-Specific; WS = worksheet; SD = standard deviation; Low Patient = one standard deviation below the mean of patient homework (between-session) worksheet competence; High Patient = one standard deviation above the mean of patient homework worksheet competence. The shaded areas indicate the levels of worksheet competence at which the impact of homework worksheet competence on change in PTSD was statistically significant.

2.18, $p = .03$, $d = .48$). In contrast to the significant Session 2 patient competence main effect for this time period, the Session 3 effect was positive, indicating that higher Session 3 patient competence scores were

associated with smaller decreases in PTSD from Session 8 to Session 10.

We found some support for our hypothesis that the combination of high in-session and out-of-

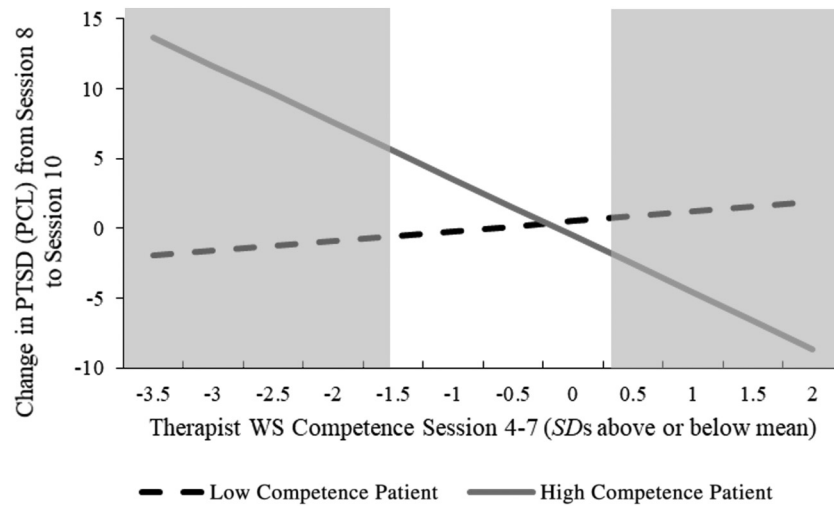


FIGURE 3 The impact of therapist-guided, in-session worksheet competency (Sessions 4-7) on change in PTSD from Session 8 to Session 10 as a function of patient worksheet competency (Sessions 4-7). PTSD = Posttraumatic Stress Disorder; PCL = PTSD Checklist – Stressor-Specific; WS = worksheet; SD = standard deviation; Low Patient = one standard deviation below the mean of homework worksheet competence; High Patient = one standard deviation above the mean of in-session worksheet competence. The shaded areas indicate the levels of worksheet competence at which the impact homework worksheet competence on change in PTSD was statistically significant.

session homework competence would be associated with greater symptom change. The main effect of patient homework competence assigned at Session 2 was qualified by a significant patient by in-session worksheet competence interaction ($b = -11.54$, $cr = -3.69$, $p < .001$). The effect size of the interaction was large ($d = -.80$). Figure 2 illustrates the nature of these interactions. For patients with *high* levels of competence as assessed on completed between-session worksheets, in-session worksheet competence exhibited a negative association with change between Session 8 and 10 and between Session 12 and posttreatment, such that higher in-session worksheet competence was associated with larger symptom decreases relative to lower in-session worksheet competence for those more skilled patients. Individuals with *low* levels of between-session worksheet competence experienced less symptom improvement if they and their therapist completed in-session worksheets with higher competence, compared to those for whom in-session worksheets were completed with lower levels of competence. For the Sessions 4-7 worksheet competence scores, a similar, significant patient-by-therapist interaction emerged for change in PTSD between Session 8 and Session 10, although as Figure 3 suggests, the difference was less pronounced for patients with lower levels of between-session worksheet competence than those with higher levels of competence.

Discussion

In this proof of concept study, we assessed whether cognitive therapy worksheets completed in CPT sessions and evaluated by external raters could be used to assess fidelity by examining the feasibility and acceptability, interrater reliability, and criterion-related validity of this method. We also explored whether the quality of worksheets was associated with subsequent symptom change. Findings revealed that, if refined and supported through additional research, worksheets have the potential to offer a reliable, feasible, and efficient strategy for measuring fidelity. While not associated with overall session fidelity (which includes CPT elements that are not included in worksheets such as psychoeducation and Socratic questioning), worksheet fidelity scores are strongly associated with observers' fidelity ratings of some key CPT skills and processes that are central to symptom change according to CPT's underlying theory (e.g., identifying thoughts and feelings, cognitive restructuring). We also found that the skill with which therapists guide worksheet completion in sessions was associated with symptom reduction among patients who were more

skilled, as evidenced by correctly completing the worksheets with minimum therapist input.

These findings have potentially important implications for implementation research and practice. Ongoing fidelity monitoring and support has been shown to reduce turnover among trained therapists (Aarons et al., 2009), and higher fidelity is associated with greater PTSD symptom reduction (Farmer et al., 2017; Marques et al., 2019). Coding worksheets for fidelity is a strategy that may be feasible for assessing the use of specific cognitive components and the quality with which they are implemented, without time- and resource-intensive observation of full sessions or the subjectivity that accompanies self-reports of fidelity. The ability of raters with varying degrees of experience with CPT to achieve good rater agreement suggests that this method could be feasible for use in clinical care settings, although if additional research and modification will be necessary to maximize the clinical utility of this novel approach.

It is important to note that although worksheet ratings may be useful in some ways, we did not find evidence that this strategy reflects every element of CPT that observer-rated fidelity captures. The worksheet adherence and competence ratings were not significantly correlated with observer ratings of overall CPT fidelity based on videos of therapy sessions, but they were highly correlated with observer ratings of the CPT session skill and process-based elements that are addressed in the worksheets. Thus, the use of worksheets cannot necessarily be used as a substitute for overall CPT fidelity. However, the link between worksheet competence and symptom change indicates that they can be used to assess aspects of the delivery of some key cognitive interventions. In previous research, the type of cognitive change that worksheets target has been associated with symptom change (Schumm et al., 2015). Whether this cognitive change arises uniquely from in-session behaviors such as Socratic dialogue, which could not be captured through worksheet ratings, or from doing worksheets remains to be investigated. Worksheets that are completed independently between sessions can also be used to assess treatment receipt, as reflecting the quality of patients' independently completed worksheets. Whereas the time required to observe a session and rate fidelity is at least as long as the session itself, the strategy of evaluating worksheets required, on average, only approximately 7 minutes. Even if videos can be viewed at 1.25 or 1.50 speed, full session observer ratings can require that the entire 50+ minute session be viewed and then time taken to complete ratings. Additionally, in contrast

to the added effort of digitally recording therapy sessions and transmitting them to expert therapists for review, worksheets are routinely completed and saved in CPT and thus may be more convenient and acceptable to therapists and their patients. If further research supports this approach, it could have important implications for mental health systems and organizations that cannot feasibly assess fidelity of some key elements using observational strategies. The approach evaluated in this study may be a promising strategy for evaluating cognitive therapy fidelity and has the potential to reduce the burden on raters in terms of both time and cost in clinical trials as well as routine clinical practice.

Although we did not find evidence for a direct association between ratings of worksheets completed in session and symptom change, the finding that worksheet-based therapist competence was associated with symptom change with more skilled patients suggests that the work therapists do in session—such as initially teaching the skills of using the worksheets, reviewing worksheets, and, when necessary, providing corrective feedback—interacts with the patient's ability to complete the worksheets independently. The quality of the worksheets completed in session with the therapist and the worksheets completed outside of session were significantly correlated in many sessions, suggesting that in-session worksheet competence may be influenced by patient abilities independent of therapist skill. Importantly, our analytic strategy took this relationship into account. It is possible that there are floor effects in later sessions for some patients who demonstrate high levels of independent skill and symptom change early in treatment, which may contribute to attenuated symptom change in later sessions. It is also possible that these interactions reflect the positive effect of alignment between therapist and patients competence, whereby focus on perfecting a worksheet at the expense of other in-session elements that may facilitate change may be inappropriate for patients who struggle to use the skills independently. Such findings may be expected in complex interventions that involve interactions between and contributions from both the provider and the patient, and our coding method can be used to facilitate future exploration of these interactions.

While there remains a need for additional research on this method, our findings highlight the importance of assessing both the quality of work done with the guidance of the therapist in session and work done independently outside of the session, or treatment receipt (Toomey et al., 2015). It can be challenging to disentangle the contributions of therapists and pa-

tients to clinical outcomes using observational methods (Webb et al., 2010). Given that the current study suggests that both patient and therapist skills may be needed to maximize outcomes, assessments of both are important. Of course, these assessments are not entirely independent. Like observer-rated fidelity, ratings of worksheets completed in session are not a pure measure of therapist skill, particularly if the patients are the ones who actually write on the worksheets in session. It may be difficult for the therapist to ensure that patients write down the responses to worksheet prompts exactly as articulated in the session, and therapists may choose not to correct every element of a worksheet in the session if it could impact the therapeutic alliance or the ability to move forward with new material. Future investigation of relationships between therapist competence, patient competence, and clinical change could benefit from asking therapists to write on at least one worksheet in session as their patients observe.

Although the findings of this study suggest that a worksheet-based approach to fidelity assessment holds promise, there are some limitations that are important to bear in mind. Because therapist skill levels were not experimentally manipulated, our findings suggest associations, but cannot confirm the impact of therapist competence on PTSD symptom change. In addition, the worksheets collected for this study were completed in the context of treatment in a military behavioral health clinic for the purpose of a randomized clinical trial. Additionally, this study was conducted with an active duty military sample that was predominantly male, although the study participants were diverse in terms of race, ethnicity, and education levels. Future research in routine care settings can be used to determine whether these findings extend to a wider variety of patients, therapists, and worksheet quality (Wiltsey Stirman, Marques, et al., 2018). Although some of the study therapists were newly trained in CPT, the findings may not generalize to CPT worksheets completed in routine clinical care settings. Worksheets were generally completed with at least a moderate level of competence. The Cronbach's α were also acceptable but somewhat low (.60-.68) for homework and in-session worksheet competence. However, these results may reflect that Cronbach's alphas tend to be lower for scales with fewer items. Furthermore, as Streiner (2003) notes, the adherence and competence assessments may be more appropriately considered an index rather than a scale, because they comprise items that may not be correlated (e.g., skill at differentiating thoughts from feelings and identifying rational responses) but that assess different aspects of fidelity. Finally, future research

efforts that involve the assessment of symptoms at every session, rather than every other session, can allow a more precise determination of which worksheets or sessions are most closely related to symptom change.

Despite these limitations, this study has potential implications for advancing the measurement of fidelity as well as interactions between patient and therapist skills on essential cognitive elements of CPT. While more research to compare this approach to gold-standard observer ratings is necessary, our findings suggest that worksheet ratings may facilitate an examination of the complex and nuanced relationships between therapist and patient competence as they relate to treatment outcomes. Efforts to scale up delivery of CBTs and to monitor fidelity can be limited by availability of personnel and the time required for observation. Identifying a strategy that is less time intensive has implications for efforts to assess fidelity, whether fidelity is evaluated as an implementation outcome, a process variable in treatment research, a tool for ongoing feedback, or as a criterion for successful completion of treatment. Current findings did not establish a direct link between in-session worksheet quality and symptom change. Further investigation is warranted before any high-stakes decisions (e.g., quality benchmarks, substitution of worksheets ratings for observer ratings for therapist certification) are made based on worksheets, but additional refinement and research on this approach may result in a scalable alternative to observer ratings. Research is currently being conducted to determine whether a refined approach works for CPT fidelity assessment, as well as CBTs for other disorders (Wiltsey Stirman, Marques, et al., 2018). If further research demonstrates the validity of this alternative method, the use of worksheets to assess fidelity has the potential to greatly increase the scalability and feasibility of large-scale fidelity assessment of CBTs in clinical care.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest.

References

- Aarons, G., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*, 38, 4–23. <https://doi.org/10.1007/s10488-010-0327-7>
- Aarons, G. A., Sommerfeld, D. H., Hecht, D. B., Silovsky, J. F., & Chaffin, M. J. (2009). The impact of evidence-based practice implementation and fidelity monitoring on staff turnover: Evidence for a protective effect. *Journal of Consulting and Clinical Psychology*, 77, 270–280. <https://doi.org/10.1037/a0013223>
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Bass, J. K., Annan, J., McIvor Murray, S., Kaysen, D., Griffiths, S., Cetinoglu, T., et al. (2013). Controlled trial of psychotherapy for Congolese survivors of sexual violence. *New England Journal of Medicine*, 368, 2182–2191. <https://doi.org/10.1056/NEJMoa1211853>
- Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., et al. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology*, 73, 852–860. <https://doi.org/10.1037/0022-006X.73.5.852>
- Brosnan, L., Reynolds, S., & Moore, R. G. (2008). Self evaluation of cognitive therapy performance: Do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy*, 36, 581–587. <https://doi.org/10.1017/S1352465808004438>
- Carroll, K. M., Martino, S., & Rounsaville, B. J. (2010). No train, no gain? *Clinical Psychology: Science and Practice*, 17, 36–40. <https://doi.org/10.1111/j.1468-2850.2009.01190.x>
- Chambers, D. (2012, December). Sustainability. *Paper presented at the Training Institute for Dissemination and Implementation Research In Health, San Jose, CA.*
- Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for normed and standardized instruments in psychology. *Psychological Assessment*, 6, 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, 23, 318–327. <https://doi.org/10.3109/09540261.2011.606803>
- Conklin, L. R., & Strunk, D. R. (2015). A session-to-session examination of homework engagement in cognitive therapy for depression: Do patients experience immediate benefits? *Behavior Research and Therapy*, 72, 56–62. <https://doi.org/10.1016/j.brat.2015.06.011>
- Creed, T. A., Wiltsey Stirman, S. W., Evans, A. C., & Beck, A. T. (2014). A model for implementation of cognitive therapy in community mental health: The Beck Initiative. *The Behavior Therapist*, 37(3), 56–64.
- Decker, S., Carroll, K. M., Nich, C., Canning-Ball, M., & Martino, S. (2013). Correspondence of motivational interviewing adherence and competence ratings in real and role-played client sessions. *Psychological Assessment*, 25, 306–312. <https://doi.org/10.1037/a0030815>
- Dusenbury, L., Brannigan, R., Falco, M., & Lake, A. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18, 237–256. <https://doi.org/10.1093/her/18.2.237>
- Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy*, 49, 373–378. <https://doi.org/10.1016/j.brat.2011.03.005>
- Farmer, C. C., Mitchell, K. S., Parker-Guilbert, K., & Galovski, T. E. (2017). Fidelity to the cognitive processing therapy protocol: Evaluation of critical elements. *Behavior Therapy*, 48, 195–206. <https://doi.org/10.1016/j.brat.2011.03.005>
- Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of*

- Consulting and Clinical Psychology*, 67, 578–582. <https://doi.org/10.1037/0022-006X.67.4.578>
- Feldstein, A. C., Glasgow, R. E., & Smith, D. H. (2008). A Practical, Robust Implementation and Sustainability Model (PRISM) for integrating research findings into practice. *Joint Commission Journal on Quality and Patient Safety*, 34, 228–243. <https://doi.org/10.2165/00148365-200806040-00001>
- Galovski, T. E., Blain, L. M., Chappuis, C., & Fletcher, T. (2013). Sex differences in recovery from PTSD in male and female interpersonal assault survivors. *Behaviour Research and Therapy*, 51, 247–255. <https://doi.org/10.1016/j.brat.2013.02.002>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20, 102–116. <https://doi.org/10.1037/a0038889>
- Hrisos, S., Eccles, M., Francis, J., Dickinson, H., Kaner, E., Beyer, F., et al. (2009). Are there valid proxy measures of clinical behaviour? A systematic review. *Implementation Science*, 4, 37. <https://doi.org/10.1186/1748-5908-4-37>
- Jensen-Doss, A., Cusack, K. J., & de Arellano, M. (2008). Workshop-based training in trauma-focused CBT: An in-depth analysis of impact on provider practices. *Community Mental Health Journal*, 44, 227–244. <https://doi.org/10.1007/s10597-007-9121-8>
- Johnson-Kozlow, M., Hovell, M. F., Rovniak, L. S., Sirikulvadhana, L., Wahlgren, D. R., & Zakarian, J. M. (2008). Fidelity issues in secondhand smoking interventions for children. *Nicotine & Tobacco Research*, 10, 1677–1690. <https://doi.org/10.1080/14622200802443429>
- Karlin, B. E., Brown, G. K., Trockel, M., Cuning, D., Zeiss, A. M., & Taylor, C. B. (2012). National dissemination of cognitive behavioral therapy for depression in the Department of Veterans Affairs Health Care System: Therapist and patient level outcomes. *Journal of Consulting & Clinical Psychology*, 80, 707–718. <https://doi.org/10.1037/a0029328>
- Karlin, B., & Cross, G. (2014). From the laboratory to the therapy room: National dissemination and implementation of evidence-based psychotherapies in the US Department of Veterans Affairs Health Care System. *American Psychologist*, 69, 19–33. <https://doi.org/10.1037/a0033888>
- Kauth, M. R., Sullivan, G., Blevins, D., Cully, J. A., Landes, R. D., Said, Q., et al. (2010). Employing external facilitation to implement cognitive behavioral therapy in VA clinics: A pilot study. *Implementation Science*, 5, 75. <https://doi.org/10.1186/1748-5908-5-75>
- Liddle, H. A., Rowe, C., Gonzalez, A., Henderson, C. E., Dakof, G., & Greenbaum, P. (2006). Changing provider practices, program environment, and improving outcomes by transporting multidimensional family therapy to an adolescent drug treatment setting. *The American Journal on Addictions*, 15, 102–112. <https://doi.org/10.1080/10550490601003698>
- Marques, L., Valentine, S. E., Kaysen, D., Mackintosh, M. -A., De Silva, D., Louise, E., et al. (2019). Provider fidelity and modifications to cognitive processing therapy in a diverse community health clinic: Associations with clinical change. *Journal of Consulting and Clinical Psychology*, 87, 357–369. <https://doi.org/10.1037/ccp0000384>
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577–605. <https://doi.org/10.1146/annurev.psych.60.110707.163612>
- Monson, C. M., Schnurr, P. P., Resick, P. A., Friedman, M. J., Young-Xu, Y., & Stevens, S. P. (2006). Cognitive processing therapy for veterans with military-related posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, 74, 898–907. <https://doi.org/10.1037/0022-006X.74.5.898>
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Nakamura, B. J., Chorpita, B. F., Hirsch, M., Daleiden, E., Slavin, L., Amundson, M., et al. (2011). Large-scale implementation of evidence-based treatments for children 10 years later: Hawaii's evidence-based services initiative in children's mental health. *Clinical Psychology: Science and Practice*, 18, 24–35. <https://doi.org/10.1111/j.1468-2850.2010.01231.x>
- Peterson, A. L., Roache, J. D., Raj, J., & Young-McCaughan, S., for the STRONG STAR Consortium (2013). The need for expanded monitoring of adverse events in behavioral health clinical trials. *Contemporary Clinical Trials*, 34, 152–154. <https://doi.org/10.1016/j.cct.2012.10.009>
- Proctor, E., Silmere, H., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., et al. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38, 65–76. <https://doi.org/10.1007/s10488-010-0319-7>
- Resick, P. A., Galovski, T. E., Uhlmansiek, M. O. B., Scher, C. D., Clum, G. A., & Young-Xu, Y. (2008). A randomized clinical trial to dismantle components of cognitive processing therapy for posttraumatic stress disorder in female victims of interpersonal violence. *Journal of Consulting and Clinical Psychology*, 76, 243–258. <https://doi.org/10.1037/0022-006X.76.2.243>
- Resick, P. A., Nishith, P., Weaver, T. L., Astin, M. C., & Feuer, C. A. (2002). A comparison of cognitive processing therapy, prolonged exposure and a waiting condition for the treatment of posttraumatic stress disorder in female rape victims. *Journal of Consulting and Clinical Psychology*, 70, 867–879. <https://doi.org/10.1037/0022-006X.70.4.867>
- Resick, P. A., & Schnicke, M. K. (1992). Cognitive processing therapy for sexual assault victims. *Journal of Consulting and Clinical Psychology*, 60, 748–756. <https://doi.org/10.1037/0022-006X.60.5.748>
- Resick, P. A., Wachen, J. S., Dondanville, K. A., Pruiksma, K. E., Yarvis, J. S., Peterson, A. L. et al.; and the STRONG STAR Consortium. (2017). Effect of group vs individual cognitive processing therapy in active-duty military seeking treatment for posttraumatic stress disorder: A randomized clinical trial. *JAMA Psychiatry*, 74, 28–36. <https://doi.org/10.1001/jamapsychiatry.2016.2729>
- Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, 38, 32–43. <https://doi.org/10.1007/s10488-010-0321-0>
- Schumm, J. A., Dickstein, B. D., Walter, K. H., Owens, G. P., & Chard, K. M. (2015). Changes in posttraumatic cognitions predict changes in posttraumatic stress disorder symptoms during cognitive processing therapy. *Journal of Consulting and Clinical Psychology*, 83, 1161–1166. <https://doi.org/10.1037/ccp0000040>
- Shediac-Rizkallah, M. C., & Bone, L. R. (1998). Planning for the sustainability of community-based health programs: conceptual frameworks and future directions for research, practice and policy. *Health Education Research*, 13, 87–108. <https://doi.org/10.1093/her/13.1.87>
- Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, 80, 217–222. https://doi.org/10.1207/S15327752JPA8003_01

- Strunk, D. R., Brotman, M. A., DeRubeis, R. J., & Hollon, S. D. (2010). Therapist competence in cognitive therapy for depression: Predicting subsequent symptom change. *Journal of Consulting and Clinical Psychology, 78*, 429–437. <https://doi.org/10.1037/a0019631>
- Toomey, E., Currie-Murphy, L., Matthews, J., & Hurley, D. A. (2015). Implementation fidelity of physiotherapist-delivered group education and exercise interventions to promote self-management in people with osteoarthritis and chronic low back pain: A rapid review part II. *Manual Therapy, 20*, 287–294. <https://doi.org/10.1016/j.math.2014.10.012>
- Vogt, D., Shipherd, J. C., & Resick, P. A. (2012). Posttraumatic maladaptive beliefs scale: Evolution of the personal beliefs and reactions scale. *Assessment, 19*, 308–317. <https://doi.org/10.1177/10731911110376161>
- Ward, A. M., Regan, J., Chorpita, B. F., Starace, N., Rodriguez, A., Okamura, K. et al.; and The Research Network on Youth Mental Health (2013). Tracking evidence based practice with youth: Validity of the MATCH and Standard Manual Consultation Records. *Journal of Clinical Child & Adolescent Psychology, 42*, 44–55. <https://doi.org/10.1080/15374416.2012.700505>
- Weathers, F. W., Litz, B. T., Huska, J. A., & Keane, T. M. (1993, October). The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility. *Paper presented at the annual meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.*
- Webb, C. A., Derubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*, 200–211. <https://doi.org/10.1037/a0018912>
- Wiltsey Stirman, S., Gutner, C. A., Suvak, M. K., Adler, A., Calloway, A., & Resick, P. (2018). Homework completion, patient characteristics, and symptom change in cognitive processing therapy for PTSD. *Behavior Therapy, 49*, 741–755. <https://doi.org/10.1016/j.beth.2017.12.001>
- Wiltsey Stirman, S., Marques, L., Creed, T. A., Gutner, C. A., DeRubeis, R., Barnett, P. G., et al. (2018). Leveraging routine clinical materials and mobile technology to assess CBT fidelity: The Innovative Methods to Assess Psychotherapy Practices (imAPP) study. *Implementation Science, 13*, 69. <https://doi.org/10.1186/s13012-018-0756-3>
- Wiltsey Stirman, S., Pontoski, K., Creed, T., Xhezo, R., Evans, A. C., Beck, A. T., et al. (2015). A non-randomized comparison of strategies for consultation in a community-academic training program to implement an evidence-based psychotherapy. *Administration and Policy in Mental Health and Mental Health Services Research, 44*, 55–66. <https://doi.org/10.1007/s10488-015-0700-7>

RECEIVED: February 26, 2020

ACCEPTED: August 28, 2020

AVAILABLE ONLINE: 12 September 2020