Measuring Memory for Treatment Using Patient Conceptualizations of Clinical Vignettes: A Pilot Psychometric Study in the Context of Cognitive Therapy for Depression

Garret G. Zieve¹ · Courtney C. Armstrong¹ · Ian M. Richardson¹ · Sydney B. Garcia¹ · Allison G. Harvey¹

Accepted: 24 January 2022 / Published online: 1 February 2022
© The Author(s) 2022

Abstract

Background Patient memory for psychological treatment contents is a promising transdiagnostic mechanism of change, but there is little consensus concerning its measurement. We conducted a pilot psychometric investigation of the Conceptualization Task, a novel measure of patient memory for treatment.

Methods Data were from a trial comparing cognitive therapy-as-usual to cognitive therapy plus the Memory Support Intervention (MSI) for adults with depression ($N = 171$). For the Conceptualization Task, patients read clinical vignettes and provided written responses to assess three facets of conceptualization: identifying contributing factors to psychopathology, making intervention recommendations, and providing a rationale for recommendations. Higher scores were given to responses reflecting accurate memory for the theoretical model and change strategies used in treatment.

Results The Conceptualization Task showed excellent inter-rater reliability and sensitivity to change during treatment, but only fair test–retest reliability and insufficient internal consistency. Findings supported discriminant validity with measures of education, IQ, and general memory functioning, but not convergent validity with existing measures of patient memory for treatment. Criterion validity analyses showed that some aspects of the Conceptualization Task were associated with therapist use of memory support strategies from the MSI and treatment outcome. However, findings were mixed, effect sizes were small, and some results did not remain statistically significant after correcting for multiple comparisons.

Conclusions Further refinement and testing is needed before the Conceptualization Task may be used to assess the patient memory for treatment contents.

Keywords Patient recall · Memory support · Depression · Cognitive therapy

Abbreviations

CT	Cognitive therapy
IDS-SR	Inventory of depressive symptoms-self report
MSI	Memory support intervention
MSRS	Memory Support Rating Scale
WHODAS	WHO Disability Assessment Scale

An emerging area of research has identified patient memory for psychological treatment contents as a promising transdiagnostic mechanism of change (Harvey et al., 2014; Zieve

et al., 2019a). This work is grounded on several lines of evidence. First, studies of a variety of treatments ranging from behavioral treatment of insomnia to marital communication skills training suggest that patient memory for treatment contents is limited (Chambers, 1991; Gumport et al., 2015; Hahlweg & Richter, 2010; Lee & Harvey, 2015). Second, worse memory for treatment contents is associated with worse treatment adherence and worse treatment outcome (Dong et al., 2017; Lee & Harvey, 2015; Zieve et al., 2020). *Treatment contents* are defined as the insights, skills, and strategies discussed during therapy that the treatment protocol specifies are important for the patient to remember and/or implement as a part of treatment (Harvey et al., 2014).

Building on these initial findings, a third line of evidence suggests that improving patient memory for treatment contents with the Memory Support Intervention (MSI) can improve treatment outcome (Harvey et al., 2016). The MSI is composed of eight empirically derived strategies that

✉ Allison G. Harvey
aharvey@berkeley.edu

¹ Department of Psychology, University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA 94720-1650, USA

therapists incorporate intensively into treatment-as-usual (Table 1; Harvey et al., 2014). Although treatment-as-usual may involve some memory support strategies (e.g., asking the patient to summarize at the end of the session), the MSI involves incorporating memory support strategies more frequently relative to treatment-as-usual. Importantly, the goal of the MSI is to increase patient memory for treatment contents specifically, rather than to enhance general memory functioning per se, in order to improve outcome.

Notably, some memory support strategies may be more effective at promoting patient memory for treatment contents than others. In educational psychology, activities involving the construction of new ideas, inferences, or connections, which go beyond what is explicitly presented by instructional materials, result in better memory outcomes than activities involving the passive absorption of information (e.g., listening to a lecture) or interacting with learning material without generating new content (e.g., rehearsing previously learned information) (Chi & Wylie, 2014; Menekse et al., 2013). Regarding the MSI, evidence suggests that memory support strategies involving the construction of new ideas, inferences, or connections about treatment contents (“constructive memory support strategies”) are more

strongly associated with patient memory for treatment than non-constructive memory support strategies (Zieve et al., 2019b). Constructive memory support strategies in the MSI include categorization, evaluation, application, and cue-based reminder, whereas non-constructive strategies include attention recruitment, repetition, practice remembering, and praise recall (Table 1). To illustrate, practice remembering is classified as non-constructive because it involves the patient restating previously discussed treatment contents, but not generating new ideas, inferences, or connections. In contrast, application is classified as constructive because it involves the patient generating new ideas regarding the use of treatment contents in a novel scenario.

A central question concerning this emerging literature is how to optimally measure patient memory for treatment contents. An answer involves identifying not only which aspects of patient memory for treatment are most relevant to promoting positive treatment outcomes but also what format should be used to measure them. Converging on a gold-standard approach to assessing patient memory for treatment may help refine the content and focus of the MSI.

Most studies examining patient memory for treatment contents have used free recall tasks (Chambers, 1991;

Table 1 Memory support strategies

Strategy	Constructive/ non-constructive [†]	Definition
Attention Recruitment	Non-constructive	Involves the therapist using expressive language that explicitly communicates to the patient that a treatment point is important to remember (e.g., “if there is one thing I would like you to remember in 10 years time, it is this skill”), or multimedia/diverse presentation modes (e.g., using a white board) as a means to recruit the patient’s attention
Repetition	Non-constructive	Involves the therapist restating, rephrasing, or revisiting information discussed in treatment (e.g., “in other words,” “as we talked about earlier,” or “in sum”)
Practice Remembering	Non-constructive	Involves the therapist facilitating the patient to regenerate, restate, rephrase, and/or revisit a treatment point (e.g., “Can you tell me some of the main ideas you’ve taken away from today’s session?”)
Praise Recall	Non-constructive	Involves the therapist rewarding the patient for successfully recalling a treatment point (e.g., “It’s really great that you remembered that point!”) or remembering to implement a desired treatment point (e.g., “I’m so glad you remembered to step back and look at the evidence.”)
Categorization	Constructive	Involves explicit effort by the therapist to work with the patient to group treatment points discussed into common themes/principles (e.g., “Let’s create a list of ways we can work on waking up at the same time each morning.”)
Evaluation	Constructive	Involves the therapist working with the patient to (a) discuss the pros/cons of a treatment point (e.g., “What would be some advantages/disadvantages of waking up at the same time each morning?”); or (b) use comparisons to compare a new treatment point to an existing or hypothetical alternative (e.g., “How would this new strategy of exercising more compare to your current habit of lying in bed when you are feeling depressed?”)
Application	Constructive	Involves the therapist working with the patient to apply a treatment point to past, present, or future (real or hypothesized) scenarios (e.g., “Can you think of an example in which you might try this new method of coping to deal with your stress at work?”)
Cue-Based Reminder	Constructive	Involves the therapist helping the patient develop new or existing cues (e.g., text reminders) to facilitate memory for treatment points

[†]Constructive memory support strategies involve patients generating new ideas, inferences, or connections about treatment contents that go beyond what has been explicitly presented by the therapist (Chi & Wylie, 2014). Non-constructive memory support strategies do not involve patients generating new ideas, inferences, or connections

Hahlweg & Richter, 2010; Harvey et al., 2016; Lee & Harvey, 2015). Among the free recall tasks used across studies, the Patient Recall Task has the best established reliability and validity (Harvey et al., 2016; Lee & Harvey, 2015). For this task, patients are given 10 min to write down all the treatment points they can remember from their sessions. *Treatment points* are defined as the individual insights, skills, and strategies discussed during therapy that together comprise the *treatment contents*. Additional methods used to measure patient memory for treatment include the Thoughts and Application Task, which involves asking patients to describe recent instances when they thought about or applied treatment points between sessions, and the Generalization Task, which involves presenting patients with hypothetical scenarios relevant to their presenting problem and asking them to describe what they would think and do (Gumport et al., 2015, 2018). Some studies have also used quiz formats including true/false and multiple choice questions about treatment points (e.g., true/false: viewing yourself as a bad person will help you to correct your behavior) (Andersson et al., 2012; Berg et al., 2019; Kronmüller et al., 2007).

These approaches to measuring patient memory for treatment each have strengths and weaknesses. The free-recall format of the Patient Recall Task replicates recall conditions patients face when needing to remember treatment contents in daily life. However, this task assesses only declarative memory and may not reflect memory that occurs on an implicit or procedural level (Schacter, 1987). In contrast to the Patient Recall Task, the Thoughts and Application Task and Generalization Task assess recall of treatment points in applied scenarios. However, by focusing on applied scenarios, these tasks may not adequately assess memory for treatment points that are more conceptual (e.g., the underlying theoretical model used in treatment). Quiz formats allow for the assessment of baseline knowledge about treatment contents, whereas the framing of the Patient Recall Task, Thoughts and Application Task, and Generalization Task require that some sessions have occurred. However, quizzes used in past research have suffered from ceiling effects, as correct answers may be easy to infer (e.g., Andersson et al., 2012; Berg et al., 2019). Moreover, some studies have not found any association between treatment outcome and treatment quizzes (Berg et al., 2019), the Generalization Task (Gumport et al., 2018), the Thoughts and Application Task (Gumport et al., 2015), and free recall measures similar to the Patient Recall Task (Chambers, 1991; Hahlweg & Richter, 2010). Refining the measurement of patient memory for treatment contents is a continuing goal.

One unexplored and as yet unmeasured aspect of patient memory for treatment is the ability to conceptualize psychological problems in a way that reflects accurate memory for the theoretical model and change strategies used in treatment. Herein, conceptualization ability was considered to

include three facets: identifying contributing factors to psychopathology, making intervention recommendations, and providing a rationale for intervention recommendations. Such conceptualization skills are considered essential for therapists (Ellis, 2007; Persons, 2008) but may also be relevant to patients. Patients who are better able to identify contributing factors to their symptoms and select intervention strategies to target those symptoms may be more likely to improve during treatment and maintain gains after treatment.

The present study examined patient conceptualization ability using a newly developed measure in a confirmatory efficacy trial comparing cognitive therapy-as-usual (CT-as-usual) to cognitive therapy plus the MSI (CT + MSI) among adults with Major Depressive Disorder. Patient conceptualization ability was measured using patient responses to clinical vignettes, a method adopted from the therapist training literature (Eells et al., 2005; Muse & McManus, 2013; Myles & Milne, 2004). This measure is termed the Conceptualization Task. The Conceptualization Task addresses several limitations of the Patient Recall Task by assessing patient ability to recall ideas learned in treatment in applied scenarios and not relying exclusively on declarative memory, given that strong memory for the theoretical model learned in treatment can be demonstrated even if patients do not explicitly recall and explain the model. The Conceptualization Task represents a novel approach relative to the Thoughts and Application Task and Generalization Task as it includes greater assessment of patient memory for the theoretical model used in treatment, rather than a more exclusive focus on thoughts and application of the change strategies learned in treatment (e.g., cognitive restructuring). Additionally, similar to quiz formats for measuring memory for treatment, the Conceptualization Task can be used to assess baseline conceptualization ability, as the framing of the task does not require that any sessions have occurred.

This study was a pilot investigation of the Conceptualization Task, focused more on “learning than confirming” (Lee et al., 2014), and included five aims. Aim one was to evaluate the reliability of the Conceptualization Task, including inter-rater reliability, internal consistency, and test-retest reliability. Aim two was to assess the Conceptualization Task’s sensitivity to change (e.g., do scores increase following treatment?). Aim three was to investigate the construct validity of the Conceptualization Task, including convergent validity with existing measures of patient memory for treatment and discriminant validity with measures of more theoretically distant constructs (e.g., education, IQ, general declarative and working memory functioning). Aim four was to examine the criterion validity of the Conceptualization Task by testing whether higher therapist use of memory support was associated with higher Conceptualization Task scores. Analyses for this aim included both categorical measures of memory support (e.g., treatment condition:

CT-as-usual vs. CT + MSI) and continuous measures of memory support (e.g., the number of constructive and non-constructive memory support strategies used per session). Aim five was to further examine the criterion validity of the Conceptualization Task by testing whether higher Conceptualization Task scores were associated with better treatment outcome on measures of depression and functional impairment.

Method

Study Overview

Data for the current study were drawn from a confirmatory efficacy trial (R01MH108657, ClinicalTrials.gov identifier: NCT02938559) comparing cognitive therapy-as-usual (CT-as-usual) to cognitive therapy integrated with the MSI (CT + MSI) in a sample of 178 adults with Major Depressive Disorder (MDD). MDD diagnoses were established using the Structured Clinical Interview for DSM-5-Research Version (First et al., 2015). Patients in both conditions received 20–26 weekly, 50-min, individual sessions over 16 weeks. Therapists in both conditions delivered cognitive therapy according to an identical protocol (Beck, 2011), using workbooks and handouts of the same quality and quantity (Greenberger & Padesky, 2015). Therapist competence in delivering cognitive therapy was confirmed using the Cognitive Therapy Rating Scale (Young & Beck, 1980). Therapists in the CT + MSI condition were additionally trained to use the strategies in the MSI and were instructed to incorporate these strategies frequently during sessions. Because several recommended practices within CT-as-usual may function as memory support strategies (e.g., providing regular capsule summaries may function as repetition), there was also some level of memory support used in the CT-as-usual condition. The mean number of memory support strategies used per session was 8.15 (*SD* 2.94) in the CT-as-usual condition, and 16.82 (*SD* 5.37) in the CT + MSI condition, according to the Memory Support Rating Scale (described below). It is important to note that therapists in the CT + MSI condition were not trained to distinguish between non-constructive and constructive memory support, as the study team was not aware of this important distinction made in the educational psychology literature at the time.

Patients completed assessments at pre-treatment, monthly timepoints during treatment, post-treatment, as well as 6- and 12-month follow-ups. The Conceptualization Task was completed only at pre-treatment, post-treatment, and at the 6-month follow-up. Thus, this report is based on analyses of data from these three timepoints. A table summarizing the assessment timepoints for each measure included in this study is available in Online Appendix A. Due to

administrative error, seven patients did not complete the Conceptualization Task at any time point and were excluded from this analysis. Demographic characteristics for the 171 patients analyzed for this study are displayed in Table 2. The University of California, Berkeley Committee for the Protection of Human Subjects approved all study procedures. Additional details about the procedures for the trial are described elsewhere (Harvey et al., 2017).

Measures

Conceptualization Task

Patient ability to conceptualize psychological problems in a way that reflects accurate memory of the theoretical model and change strategies used in treatment was assessed with the Conceptualization Task, a new measure developed for this study. Because patients in the current study were receiving cognitive therapy for depression, the Conceptualization Task focused on assessing memory for the theoretical model and change strategies used in cognitive therapy (Beck, 2011; Greenberger & Padesky, 2015). For the task, patients read a vignette describing a person with depression and provided written responses to three open-ended questions designed to assess conceptualization ability. Patients could re-read the vignette while responding to the questions. Three vignettes were developed and delivered to each patient in a random order across the assessment points for the study. The task was introduced several months into the parent trial, resulting in missing pre-treatment data for the first 22 patients.

Vignette Content Each of the three vignettes described a person with symptoms of depression in approximately 450 words (Online Appendix B). The vignette content included information about external stressors, symptoms of depression, automatic thoughts, core beliefs, maladaptive behaviors, and important developmental experiences. Each vignette was written at a 7–8th grade reading level according to the new Dale-Chall Readability Formula (Chall & Dale, 1995).

Vignette Questions After reading a vignette, patients were given 10 min to write answers to the following open-ended questions: (1) Why do you think [name] is experiencing symptoms of depression? (2) What do you think [name] could do to start feeling better? (3) How do you think your recommendations will help [name]? These questions were designed to assess three facets of patients' conceptualization ability: identifying contributing factors to depression, making intervention recommendations, and providing a rationale for recommendations.

Table 2 Patient characteristics ($n = 171$)

Characteristic	<i>n</i>	%
Gender		
Male	63	36.84
Female	104	60.82
Declined to respond	4	2.34
Race		
White	101	59.06
Asian	26	15.20
American Indian/Alaskan Native	1	0.6
Black or African American	6	3.51
Native Hawaiian or Pacific Islander	1	0.6
Mixed	34	19.88
Declined to respond	2	1.17
Ethnicity		
Hispanic/Latino	30	17.54
Not Hispanic/Latino	137	80.17
Declined to respond	4	2.34
Marital status		
Single	104	60.82
Common law partnered	8	4.68
Married	37	21.64
Divorced	14	8.19
Separated	6	3.51
Widowed	2	1.17
Employment ^a		
Caring for others	6	3.51
Full-time	62	36.26
Part-time	41	23.98
Unemployed/looking for work	21	12.28
Unemployed/disabled	2	1.17
Retired/Unemployed not looking	12	7.02
Volunteer	5	2.92
Other	17	9.94
Declined to respond	1	0.6
Highest education level		
Some high school	1	0.6
Graduated high school	3	1.75
Completed vocational school	1	0.6
Some college	54	31.58
Completed bachelor's degree	47	27.49
Completed post baccalaureate training	3	1.75
Some graduate school	11	6.43
Completed master's degree	36	21.05
Graduate training beyond masters	8	4.68
Completed doctoral degree	7	4.09
	<i>M</i>	<i>SD</i>
Age (years)	38.08	15.70
Household income (\$)	71,912.90	63,793.23

^aPatients could select multiple categories

Scoring The goal of the scoring system for the Conceptualization Task is to capture patient responses that show cognitive therapy-consistent conceptualizations of depression. Thus, answers to each question received points for (1) the number of cognitive or behavioral factors/strategies indicated, (2) the number of statements describing connections between elements of the cognitive model, and (3) the number of cognitive therapy vocabulary terms used. These aspects were derived from the scoring systems of similar measures in the literature as well as by reviewing effective elements of exemplary responses among a pilot sample of undergraduates ($n=10$), graduate students ($n=5$), and licensed psychologists ($n=2$) who completed the task.

The first point type, *number of cognitive or behavioral factors/strategies indicated*, drew inspiration from the Patient Recall Task (Lee & Harvey, 2015) and Thoughts and Application Task (Gumport et al., 2015, 2018). Examples of responses counted under this point type include “negative beliefs about herself” for question one, “do an experiment to test if her belief is true” for question two, and “provide evidence for new positive core beliefs” for question three. Additionally, patients frequently used the colloquial expression “feels that” to mean “has the thought that” (e.g., “He feels that he does not deserve what he currently has in his life”). These answers were still counted as identifying cognitive factors/strategies.

The second point type captured when patients made statements that described *connections between elements of the cognitive model*. To gain a point here, statements had to describe a connection between at least one behavior or cognitive factor and another element of the cognitive model. Examples of responses counted under this point type include: “She got fired from work which exacerbated her negative beliefs about herself” for question one (which connects environmental and cognitive factors), “create a schedule with mastery, pleasure activities, help her to feel purpose again” for question two (which connects behavioral and cognitive factors), and “as she builds action plans, her life will likely begin to improve, providing more evidence for new positive core beliefs” for question three (which connects behavioral, environmental, and cognitive factors). Responses describing connections between two behavioral factors or cognitive factors (e.g., “Identifying things she is grateful for which can help her maybe not see her situation as being so bad after all”) were also scored under this point type.

The third point type was awarded when patients used *cognitive therapy vocabulary terms*. An a priori list of vocabulary terms was constructed by reviewing therapist training materials and treatment materials used with patients (Beck, 2011; Greenberger & Padesky, 2015). Patients received one point any time they used a term from the list in their responses. Example terms from the vocabulary list include “automatic thoughts,” “behavioral activation,” and “thinking

traps.” The full list of vocabulary terms counted in this study can be found in Online Appendix C. Example scoring breakdowns across all point types for high scoring and low scoring responses to the Kira vignette are available in Online Appendix D.

Scoring Procedures and Inter-Rater Reliability The Conceptualization Task scoring team consisted of two advanced graduate students and two research assistants. To evaluate interrater reliability, 50 task responses, representing 10% of the overall number of tasks collected in total for the study, were scored independently by each member of the team.

Convergent Validity Measures

Three existing measures of patient memory for treatment were used to assess the convergent validity of the Conceptualization Task.

Patient Recall Task The Patient Recall Task is a free recall measure of patient memory for treatment contents, during which patients are given a sheet of paper and asked to take 10 min to recall as many treatment points as they can remember from all of the sessions they have had so far. Coders then determine the overall number of distinct treatment points recalled using a scoring rubric (Lee & Harvey, 2015). The task has demonstrated excellent inter-rater reliability and is significantly associated with the amount of memory support received in previous studies (Lee & Harvey, 2015; Lee et al., 2016). In this study, inter-rater reliability was good according to a two-way mixed effects, consistency, single rater intraclass correlation ($n=100$, $ICC\ 0.89$, $p<0.001$) (Koo & Li, 2016; McGraw & Wong, 1996). The Patient Recall Task, as well as the Thoughts and Application Task and Generalization Task (described below), were not administered at pre-treatment because the framing of these tasks requires that therapy sessions have occurred.

Thoughts and Application Task Thoughts and applications of treatment points between sessions were assessed with the Thoughts and Application Task (Gumport et al., 2015, 2018). For this task, patients provide open-ended responses to the following prompts: “In the past 24 h, did the session you attended last week come to mind? If yes, what came to mind? In the past 24 h, did you get to apply anything from the session you attended last week? If yes, what did you apply?” Responses to these prompts are coded for the number of treatment points about which patients report thinking and applying. The task has demonstrated inter-rater reliability and is significantly associated with the amount of memory support received in previous studies (Gumport et al., 2015, 2018). In this study, inter-rater reliability was substantial for both thoughts ($n=106$, Cohen’s $\kappa=0.81$,

$p < 0.001$) and applications of treatment points ($n = 106$, Cohen's $\kappa = 0.78$, $p < 0.001$) (McHugh, 2012).

Generalization Task Patient ability to generalize treatment points to novel scenarios was measured with the Generalization Task. For the task, patients are presented with two hypothetical scenarios (e.g., being rejected for a job, social rejection at a party) and are asked to describe in a free response format “what would you think?” (to assess *cognitive* generalization) and “what would you do?” (to assess *behavioral* generalization). Coders determine whether responses reflect successful generalization of treatment points, resulting in a score ranging from 0 to 2 for both cognitive and behavioral generalization. The task has demonstrated inter-rater reliability and is significantly associated with the amount of memory support received in previous studies (Gumport et al., 2015, 2018). In the current study, inter-rater reliability was substantial for both cognitive ($n = 214$, Cohen's $\kappa = 0.72$, $p < 0.001$) and behavioral generalization ($n = 214$, Cohen's $\kappa = 0.61$, $p < 0.001$) (McHugh, 2012).

Discriminant Validity Measures

Measures of educational attainment, full-scale IQ, general working memory functioning, and general declarative memory functioning were used to examine the discriminant validity of the Conceptualization Task. These measures were selected because the constructs they assess are more conceptually distant from the construct of memory for treatment, and thus should show weaker associations with the Conceptualization Task, than other measures of memory for treatment.

Education Educational attainment was operationalized as a dichotomous variable indicating whether patients had completed a 4-year college degree (yes/no).

National Adult Reading Test Full-scale IQ was estimated by proxy using the National Adult Reading Test (Nelson & Willison, 1991), during which patients were asked to pronounce a list of English words with irregular spellings designed to test vocabulary rather than the ability to apply regular pronunciation rules.

N-Back Working memory was measured with a three-back version of the N-Back task (Kirchner, 1958), during which patients were sequentially presented with a series of letters and asked to indicate whether the letter currently presented matched the letter that was presented three items previously. Performance on this task was operationalized as the proportion of correct hits minus the proportion of false positives (Snodgrass & Corwin, 1988).

Episodic Face-Name Learning Task Declarative memory was measured with the Episodic Face-Name Learning Task (Mander et al., 2011; Miller et al., 2008; Sperling et al., 2003). For the encoding phase of the task, patients were presented with a series of face images paired with names. For the retrieval phase of the task, patients were presented with previously viewed “old” faces, as well as never before seen “new” faces. Patients were asked to indicate whether each face was old or new, and if old, to choose among three name options: (1) the original name previously paired with that face, (2) a new name never shown during encoding, or (3) a name previously seen at encoding, but with a different face (the “lure name”). If new faces were incorrectly labeled as old (false positive), three names were still subsequently presented, but were new names never shown during encoding. Performance on this task was operationalized as the proportion of correctly recalled face-name pairs, minus the proportion of false positives, minus the proportion of lure names selected for correctly identified old faces.

Criterion Validity Measures

Memory Support Rating Scale Therapist use of memory support strategies was measured with the Memory Support Rating Scale (MSRS; Lee et al., 2016). The MSRS is an observer-rated measure that indexes the total number of times therapists use each of the eight different types of memory support included in the MSI within a session. The MSRS has demonstrated acceptable inter-rater reliability ($ICCs = 0.73–0.74$), discriminant validity with unrelated observer-rated measures of therapy sessions, and significant associations with patient memory for treatment (Lee et al., 2016). During the trial for this study, five sessions for each patient were selected for MSRS coding: session two, sessions from weeks 4, 8, and 12 of treatment, and the final session. MSRS coders were required to establish 80% or higher inter-rater agreement with an expert coder across five consecutive session recordings before advancing to independent coding of study data. The average percent agreement between each coder and the expert coder was 89%.

Four summary variables were derived from the MSRS by averaging individual session variables across all coded sessions for each patient. The first summary variable was the average total number of memory support strategies used per session. The second summary variable was the average number of *different types* of memory support strategies used per session. This was obtained by counting how many of the eight different memory support strategies therapists used at least once in a session (range 0–8). The difference between these first two summary variables can be illustrated from a hypothetical session in which a therapist used repetition three times and practice remembering two times. In this case, the total number of memory support strategies used

would be five and the number of different types of memory support strategies used would be two.

The third summary variable was the average number of constructive memory support strategies used per session. This was obtained by summing the total number of times therapists used application, evaluation, categorization, and cue-based reminders. The fourth summary variable was the average number of non-constructive memory support strategies used per session. This was obtained by summing the total number of times therapists used repetition, attention recruitment, practice remembering, and praise recall. These last two variables (average number of constructive and non-constructive memory support strategies used per session) represent novel ways of scoring the MSRS relative to the initial validation study (Lee et al., 2016). While Lee et al. found support for a one-factor structure of the MSRS, this likely reflects that training in the MSI does not yet distinguish between constructive and non-constructive memory support. Findings that constructive memory support is more strongly associated with patient memory for treatment contents than non-constructive memory support (Zieve et al., 2019b) provides criterion validity evidence that these two types of memory support are distinct constructs.

Inventory of Depressive Symptoms-Self Report Depression outcome was measured with the Inventory of Depressive Symptoms-Self Report (IDS-SR) (Rush et al., 1996), a 30-item measure of depression symptoms over the past 7 days. The IDS-SR has demonstrated high internal consistency (Cronbach's $\alpha=0.92\text{--}0.93$), convergent validity with other established measures of depression symptomatology, and sensitivity to change (Rush et al., 1996; Trivedi et al., 2004). Several dichotomous outcomes were derived from the IDS-SR following American College of Neuropsychopharmacology criteria (Rush et al., 2006): (1) *response*, defined as a 50% reduction in pre-treatment symptom severity on the IDS-SR, (2) *remission*, defined as an IDS-SR score of less than or equal to 14, and (3) *relapse*, defined as an IDS-SR score of greater than 14 at follow-up for those who had remitted by post-treatment. Remission and relapse were considered the primary outcome variables in the parent trial (Harvey et al., 2017). Response, as well as the continuous score on the IDS-SR, were also analyzed because all these outcomes were derived from the same measure.

WHO Disability Assessment Scale 2.0 Functional impairment outcome was measured with the WHO Disability Assessment Scale 2.0 (WHODAS 2.0) (Üstün et al., 2010). The WHODAS 2.0 is a 36-item self-report measure of disability in the past 30 days related to general health and mental health conditions. The WHODAS 2.0 has demonstrated high internal consistency (Cronbach's $\alpha=0.86$), test–retest reliability, convergent validity with established measures of

disability, and sensitivity to change (Üstün et al., 2010). The continuous score was considered a primary outcome variable in the parent trial (Harvey et al., 2017).

Data Analysis

For aim one, the inter-rater reliability and test–retest reliability of the Conceptualization Task were assessed using two-way mixed effects, absolute agreement, single rater intra-class correlations (ICC; Koo & Li, 2016; McGraw & Wong, 1996). ICC values were interpreted according to Cicchetti's (1994) guidelines. Test–retest reliability was evaluated using Conceptualization Task scores from post-treatment and 6-month follow-up. Pre-treatment Conceptualization Task scores were omitted from this analysis, as pre-treatment performance on the task does not reflect memory for treatment. Internal consistency was examined with Cronbach's α , using a benchmark of 0.70 for acceptable internal consistency (Cortina, 1993).

For aim two, sensitivity to change was assessed by comparing mean Conceptualization Task scores over time using repeated measures ANOVAs. For aim three, construct validity was evaluated by inspecting correlations between the Conceptualization Task and measures of convergent and discriminant validity. Correlations ≥ 0.50 were considered evidence of convergent validity, and correlations < 0.50 were considered evidence of discriminant validity (Abma et al., 2016).

For aim four, repeated measures ANOVAs were used to test whether mean Conceptualization Tasks scores differed between the CT-as-usual and CT + MSI conditions. Additionally, multiple regression was used to examine whether higher therapist use of memory support, according to continuous measures derived from the MSRS, was associated with higher Conceptualization Task scores. For aim five, multiple regression was used to test whether higher Conceptualization Task scores were associated with better treatment outcome. As the multiple regression analyses for aims four and five sought to investigate the relations between therapist use of memory support, Conceptualization Task performance, and treatment outcome regardless of treatment condition, treatment condition was tested as a covariate. As the two study conditions differed only in the dose of memory support, the relations between therapist use of memory support, Conceptualization Task scores, and treatment outcome were not expected to differ between treatment groups. Indeed, treatment condition did not show statistically significant relations with any outcomes when tested as a covariate, and thus was not included in the final analyses.

All analyses were conducted using R (R Core Team, 2016). The percentage of missing data ranged from 0 to 14.04% across measures and assessment time points. Missing data were handled using multiple imputation (Enders,

2017). As this study represented a pilot investigation, where strict significance requirements may result in low statistical power (Nakagawa, 2004), individual findings meeting an alpha criterion of 0.05 are cautiously interpreted. However, the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) was used to evaluate the robustness of findings given the multiple comparisons conducted for the study. Thus, for each aim, p -values from all analyses within each Conceptualization Task question were also compared with an adjusted significance threshold, assuming a false discovery rate of 10%.

Results

Descriptive statistics for all study variables are presented by treatment condition in Table 3.

Aim One: Reliability

Interrater reliability for individual question scores within the Conceptualization Task were in the excellent range ($ICCs = 0.89, 0.78$, and 0.85 for questions one, two, and three, respectively, all $ps < 0.001$) (Cicchetti, 1994). Internal consistency analyses were conducted to assess the appropriateness of combining the individual questions into a total score. Inter-correlations between Conceptualization Task questions can be found in Online Appendix E. Cronbach's α values were $0.46, 0.61$, and 0.57 for the pre-treatment, post-treatment, and 6-month follow-up timepoints, respectively. These values were less than the suggested cutoff of $\alpha = 0.70$ for acceptable internal consistency (Cortina, 1993). Accordingly, individual questions from the Conceptualization Task were analyzed separately instead of calculating a total score. Test-retest reliability for individual question scores were in the fair range ($ICCs = 0.42, 0.48$, and 0.48 for questions one, two, and three, respectively, all $ps < 0.001$).

Aim Two: Sensitivity to Change

Results from omnibus repeated measures ANOVAs and pairwise follow-up comparisons for Conceptualization Task scores over time are shown in Table 4. Data was pooled across treatment conditions for this analysis as Conceptualization Task scores did not differ between conditions (see Aim Four section below). Mean Conceptualization Task scores increased across all questions from pre-treatment to post-treatment, and from pre-treatment to 6-month follow-up, with small to medium effect sizes ($ds = 0.36$ – 0.77 , all $ps < 0.001$). Mean Conceptualization Task scores decreased from post-treatment to 6-month follow-up with small effect sizes for question one ($d = -0.24$, $p = 0.010$) and question two ($d = -0.17$, $p = 0.044$).

Aim Three: Construct Validity

Correlations between the Conceptualization Task and measures of convergent and discriminant validity are presented in Table 5. Regarding the convergent validity measures, the Conceptualization Task was most consistently related to the Patient Recall Task, with correlations in the small-to-medium range ($rs = 0.20$ – 0.32 , all $ps < 0.05$). Only question two of the Conceptualization Task was consistently related to the Patient Recall Task after correcting for multiple comparisons.

The Conceptualization Task was largely unrelated to the Generalization Task and Thoughts and Application Task. For the Generalization Task, only the correlations with question three on the Conceptualization Task at 6-month follow-up were statistically significant, corresponding to small effect sizes ($rs = 0.16$ – 0.18 , all $ps < 0.05$). For the Thoughts and Application Task, only the correlations with question two on the Conceptualization Task at 6-month follow-up were statistically significant, also corresponding to small effect sizes ($rs = 0.19$ – 0.21 , all $ps < 0.05$). After correcting for multiple comparisons, only the correlations between the Conceptualization Task and Thoughts and Application Task remained statistically significant.

Regarding the discriminant validity measures, only full-scale IQ was consistently uncorrelated with the Conceptualization Task. The remaining measures showed intermittent small correlations with Conceptualization Task scores. Baseline education was correlated with question two on the Conceptualization Task at pre-treatment ($r = 0.17$, $p < 0.05$). The Episodic Face-Name Learning Task was correlated with question one on the Conceptualization Task at pre-treatment ($r = 0.20$, $p < 0.05$), and with question three at post-treatment and the 6-month follow-up ($rs = 0.18$ – 0.20 , all $ps < 0.05$). Performance on the N-Back was correlated with question three on the Conceptualization Task at pre-treatment ($r = 0.17$, $p < 0.05$), and with questions one and two at post-treatment ($rs = 0.16$ – 0.25 , all $ps < 0.05$). After correcting for multiple comparisons, only the correlations between the Episodic Face-Name Learning Task and question one on the Conceptualization Task at pre-treatment, and between the N-Back and question one at post-treatment, remained statistically significant.

Aim Four: Criterion Validity Regarding Therapist Use of Memory Support

The CT-as-usual and CT + MSI conditions did not differ on Conceptualization Task performance at any timepoint, per omnibus repeated measures ANOVAs ($Fs = 0.44$ – 0.80 , $ps = 0.451$ – 0.641). Linear regression coefficients and associated effect sizes predicting Conceptualization Task performance from continuous measures of memory support

Table 3 Multiple imputation means and SDs (or *n* and % for categorical variables) for all study variables, by treatment condition

	CT-as-usual (<i>n</i> = 89)				CT + MSI (<i>n</i> = 82)			
	Pre		Post		Pre		Post	
	<i>M</i> or <i>n</i>	<i>SD</i> or %	<i>M</i> or <i>n</i>	<i>SD</i> or %	<i>M</i> or <i>n</i>	<i>SD</i> or %	<i>M</i> or <i>n</i>	<i>SD</i> or %
Conceptualization Task								
Question 1 ^a	1.95	1.86	3.42	3.11	2.02	1.73	3.41	2.95
Question 2 ^b	1.92	1.66	3.66	2.92	1.78	1.37	3.74	2.95
Question 3 ^c	1.47	1.54	2.69	2.16	1.22	1.38	2.75	2.84
Patient Recall Task	–	–	11.9	5.14	–	–	11.85	4.55
Generalization Task								
Cognitive generalization	–	–	1.02	0.85	–	–	0.89	0.80
Behavioral generalization	–	–	1.53	0.63	–	–	1.53	0.64
Thoughts and Application Task								
Thoughts	–	–	2.10	1.54	–	–	2.33	1.42
Application	–	–	1.94	1.39	–	–	1.98	1.31
Memory Support Rating Scale								
Average total MS per session	–	–	8.15	2.94	–	–	16.82	5.37
Average MS types per session	–	–	3.84	0.79	–	–	5.48	0.63
Average total non-constructive MS per session	–	–	5.64	2.10	–	–	11.78	3.79
Average total constructive MS per session	–	–	2.51	1.28	–	–	5.04	2.11
Full-Scale IQ	114.12	7.80	–	–	114.25	6.87	–	–
Episodic Face-Name Learning Task	–0.32	0.35	–0.22	0.25	–0.31	0.25	–0.27	0.35
N-Back	0.35	0.24	0.41	0.28	0.37	0.25	0.44	0.24
IDS-SR								
Total Score	39.49	7.60	18.29	12.51	38.48	7.74	17.29	11.23
Response ^d	–	–	56	62.92%	–	–	47	57.32%
Remission ^d	–	–	44	49.44%	–	–	35	42.68%
Relapse ^d	–	–	–	–	–	–	–	–
WHODAS	47.30	17.98	26.06	17.63	45.22	16.59	23.15	15.46

CT cognitive therapy, MSI memory support intervention, Pre pre-treatment, Post post-treatment, FU6 6-month follow-up, IDS-SR Inventory of Depressive Symptoms, Self-Report, WHODAS WHO Disability Assessment Scale

^aWhy do you think [name] is experiencing symptoms of depression?

^bWhat do you think [name] could do to start feeling better?

^cHow do you think your recommendations will help [name]?

^dCategorical variable

^eDenominator is number of participants who met criteria for remission at post

Table 4 Results from omnibus repeated measures ANOVAs and pairwise follow-up comparisons for Conceptualization Task scores over time ($n = 171$)

Conceptualization Task Question	<i>F</i>	<i>p</i>	Pre v. post			Pre v. FU6			Post v. FU6		
			<i>t</i>	<i>p</i>	<i>d</i>	<i>t</i>	<i>p</i>	<i>d</i>	<i>t</i>	<i>p</i>	<i>d</i>
Question 1 ^a	14.85	<0.001	5.24	<0.001	0.56	3.39	<0.001	0.36	−2.64	0.010	−0.24
Question 2 ^b	32.94	<0.001	7.19	<0.001	0.77	6.68	<0.001	0.66	2.04	0.044	−0.17
Question 3 ^c	22.26	<0.001	6.06	<0.001	0.66	5.53	<0.001	0.54	−0.77	0.455	−0.07

All statistically significant results remained significant after implementing the Benjamini and Hochberg (1995) procedure

Pre pre-treatment, *Post* post-treatment, *FU6* 6-month follow-up

^aWhy do you think [name] is experiencing symptoms of depression?

^bWhat do you think [name] could do to start feeling better?

^cHow do you think your recommendations will help [name]?

Table 5 Correlations between the Conceptualization Task and measures of convergent and discriminant validity ($n = 171$)

	Conceptualization Task Question 1			Conceptualization Task Question 2			Conceptualization Task Question 3		
	Pre	Post	FU6	Pre	Post	FU6	Pre	Post	FU6
Convergent Validity Measures									
Patient Recall Task	–	0.20*	0.30***†	–	0.22*†	0.32***†	–	0.20*	0.13
Generalization Task									
Cognitive Generalization	–	0.04	0.10	–	0.01	0.05	–	0.06	0.18*
Behavioral Generalization	–	0.11	0.06	–	0.14	0.13	–	0.03	0.16*
Thoughts and Application Task									
Thoughts	–	0.04	0.11	–	0.07	0.19*	–	0.13	0.07
Application	–	0.02	0.07	–	0.08	0.21***†	–	0.11	0.07
Discriminant Validity Measures									
Baseline Education [‡]	−0.05	−0.07	−0.02	0.17*	0.06	0.06	0.08	0.05	0.03
Baseline Full-Scale IQ	0.02	0.06	0.13	0.07	0.07	0.12	0.06	−0.03	−0.04
Episodic Face-Name Learning Task	0.20*†	0.13	0.16	0.07	0.12	0.13	−0.02	0.20*	0.18*
N-Back	0.09	0.25***†	−0.04	0.08	0.16*	0.05	0.20*	0.04	0.12

All correlations are Pearson's *r* unless otherwise noted

Conceptualization Task Question 1 Why do you think [name] is experiencing symptoms of depression?, *Question 2* What do you think [name] could do to start feeling better?, *Question 3* How do you think your recommendations will help [name]?, *Pre* pre-treatment, *Post* post-treatment, *FU6* 6-Month follow-up

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

†Remained statistically significant after implementing the Benjamini and Hochberg (1995) procedure

‡Binary variable (completed college, yes/no), point biserial correlations were used

are presented in Table 6. As the continuous measures of memory support were correlated ($r_s = 0.76$ – 0.81 , all $p_s < 0.001$), these predictors were entered into multiple regression models. The average number of memory support strategies and different types of memory support strategies used per session were entered into one model. The average number of constructive and non-constructive memory support strategies used per session were entered into a separate model. All four continuous measures of memory support could not be entered into the same model due to resulting multicollinearity (e.g., the number of constructive and non-constructive memory support strategies

together perfectly predict the total number of memory support strategies).

Conceptualization Task performance was generally not related to the number of memory support strategies and different types of memory support strategies used per session. However, a higher average number of memory support strategies used per session predicted better scores on question three of the Conceptualization Task at 6-month follow-up, corresponding to a small effect size (semi-partial $r = 0.22$, $p = 0.006$). Additionally, and contrary to study hypotheses, a greater average number of different types of memory support strategies used per session predicted worse scores on

Table 6 Linear regression coefficients and effect sizes predicting Conceptualization Task performance from continuous measures of memory support ($n = 171$)

	Post-Treatment			FU6		
	<i>b</i>	<i>p</i>	Semi-partial <i>r</i>	<i>b</i>	<i>p</i>	Semi-partial <i>r</i>
Conceptualization Task Question 1: Why do you think [name] is experiencing symptoms of depression?						
Average Total MS Per Session ^a	0.05	0.451	0.06	0.06	0.262	0.09
Average MS Types Per Session ^a	−0.09	0.823	−0.02	−0.27	0.392	−0.07
Average Total Non-Constructive MS Per Session ^b	−0.04	0.684	−0.03	−0.06	0.376	−0.07
Average Total Constructive MS Per Session ^b	0.21	0.238	0.10	0.20	0.143	0.12
Conceptualization Task Question 2: What do you think [name] could do to start feeling better?						
Average Total MS Per Session ^a	0.00	0.953	0.00	0.04	0.465	0.06
Average MS Types Per Session ^a	0.34	0.360	0.08	−0.14	0.668	−0.04
Average Total Non-Constructive MS Per Session ^b	−0.09	0.264	−0.09	−0.03	0.634	−0.04
Average Total Constructive MS Per Session ^b	0.38	0.032	0.18	0.14	0.315	0.08
Conceptualization Task Question 3: How do you think your recommendations will help [name]?						
Average Total MS Per Session ^a	0.11	0.053	0.16	0.17	0.006 [†]	0.22
Average MS Types Per Session ^a	−0.36	0.275	−0.09	−0.75	0.037 [†]	−0.18
Average Total Non-Constructive MS Per Session ^b	−0.06	0.363	−0.07	−0.03	0.707	−0.03
Average Total Constructive MS Per Session ^b	0.33	0.025 [†]	0.18	0.24	0.113	0.13

FU6 6-month follow-up, MS memory support

[†]Analysis remained statistically significant after implementing the Benjamini and Hochberg (1995) procedure. Variables marked with the same superscript were included in the same regression model

question three of the Conceptualization Task at 6-month follow-up, corresponding to a small effect size (semi-partial $r = -0.18$, $p = 0.037$). Both these results remained statistically significant after correcting for multiple comparisons.

A higher average number of constructive memory support strategies used per session predicted better scores on questions two and three of the Conceptualization Task at post-treatment, corresponding to small effect sizes (semi-partial $rs = 0.18$, all $ps < 0.05$). Conceptualization Task performance did not show any association with non-constructive memory support. After correcting for multiple comparisons, only the relation between constructive memory support strategies and question three of the Conceptualization Task remained statistically significant.

Aim Five: Criterion Validity Regarding Treatment Outcome

Regression coefficients and associated effect sizes predicting treatment outcome from Conceptualization Task performance are presented in Table 7. In addition to concurrent analyses at post-treatment and 6-month follow-up, a lagged analysis was conducted predicting 6-month follow-up treatment outcome from post-treatment Conceptualization Task performance.

Higher scores on question two of the Conceptualization Task were associated with lower total scores on the IDS-SR in the concurrent analyses at post-treatment and the 6-month

follow-up, corresponding to small effect sizes ($rs = -0.19$ to -0.17 , all $ps < 0.05$). Additionally, higher scores on question two of the Conceptualization Task were associated with lower total scores on the WHODAS in the concurrent analysis at the 6-month follow-up, corresponding to a small effect size ($r = -0.18$, $p < 0.05$). These results were not affected by adjusting for baseline IDS-SR and WHODAS scores.

To investigate whether Conceptualization Task performance was associated with independent variance in outcome, all statistically significant analyses were repeated adjusting for the existing measures of memory for treatment that were correlated with Conceptualization Task performance at each timepoint. None of the associations between Conceptualization Task performance and outcome remained statistically significant after adding these covariates. Moreover, none of the original bivariate relations between the Conceptualization Task and treatment outcome remained significant after correcting for multiple comparisons.

Discussion

This study was a pilot psychometric investigation of the Conceptualization Task in the context of cognitive therapy for depression. There was some evidence for the reliability, sensitivity to change, construct validity, and criterion validity of the Conceptualization Task. However, findings were mixed, many effect sizes were small, and some results did

Table 7 Regression coefficients and effect sizes predicting treatment outcome measures from Conceptualization Task performance ($n = 171$)

	Post-treatment (concurrent)			FU6 (concurrent)			FU6 (predicted from post-treatment)		
	b	p	Semi-partial r or d	b	p	Semi-partial r or d	b	p	Semi-partial r or d
Conceptualization Task Question 1: Why do you think [name] is experiencing symptoms of depression?									
IDS-SR									
Total Score	−0.12	0.619	−0.03	−0.57	0.153	−0.12	−0.10	0.681	−0.03
Response ^a	0.00	0.932	0.00						
Remission ^a	0.00	0.988	0.00	0.03	0.624	0.02	0.03	0.556	0.02
Relapse ^a				0.04	0.664	0.02	−0.11	0.268	−0.06
WHODAS	−0.30	0.509	−0.05	−0.78	0.172	−0.11	−0.14	0.701	−0.02
Conceptualization Task Question 2: What do you think [name] could do to start feeling better?									
IDS-SR									
Total Score	−0.69	0.035	−0.17	−0.91	0.017	−0.19	−0.15	0.586	−0.04
Response ^a	0.07	0.234	0.04						
Remission ^a	0.08	0.199	0.04	0.09	0.195	0.05	0.07	0.250	0.04
Relapse ^a				−0.05	0.596	−0.03	−0.12	0.167	−0.07
WHODAS	−0.56	0.236	−0.10	−1.22	0.036	−0.18	−0.43	0.304	−0.07
Conceptualization Task Question 3: How do you think your recommendations will help [name]?									
IDS-SR									
Total Score	−0.51	0.198	0.198	−0.24	0.52	−0.05	0.914	0.914	0.01
Response ^a	0.07	0.315	0.315						
Remission ^a	0.08	0.227	0.227	0.01	0.911	0.00	0.405	0.405	0.03
Relapse ^a				0.03	0.682	0.02	0.733	0.733	−0.02
WHODAS	−0.48	0.397	0.397	0.00	0.994	0.00	0.606	0.606	0.03

Continuous treatment outcomes were analyzed with linear regression. Categorical treatment outcomes were analyzed with logistic regression. Analyses predicting treatment outcome at FU6 from Conceptualization Task performance at post-treatment also controlled for post-treatment treatment outcome for continuous measures

IDS-SR Inventory Of Depressive Symptoms, Self-Report, WHODAS WHO Disability Assessment Scale, FU6 6-month follow-up

^aCategorical variable

not remain statistically significant after correcting for multiple comparisons. We elaborate on several complexities in study findings below.

The Conceptualization Task showed excellent inter-rater reliability, but only fair test–retest reliability. It is possible that assessing test–retest reliability from the post-treatment to 6-month follow-up timepoints allowed too much time for extraneous factors to influence Conceptualization Task scores (e.g., individual differences in forgetting treatment contents in the months following treatment). Indeed, many psychometric studies in clinical measure development use shorter intervals of one or 2 weeks to examine test–retest reliability (Streiner et al., 2015). Future research with the Conceptualization Task may consider assessing test–retest reliability over a shorter time interval.

Conceptualization Task scores increased following treatment, with small to medium effect sizes, providing some evidence that the task is sensitive to changes in patient memory for treatment. However, the parent trial for this study did not include a no-treatment control group, preventing an evaluation of practice effects that could

account for some of the increases in Conceptualization Task scores. However, using alternate versions of a test at different assessment points, as was done in this study, had been shown to attenuate practice effects (Calamia et al., 2012). Additionally, performance on several aspects of the Conceptualization Task decreased from post-treatment to 6-month follow-up, which is inconsistent with the possibility of increasing performance due to practice effects. In any case, future research should explicitly test for practice effects in Conceptualization Task performance using a no-treatment control group.

Internal consistency and convergent validity results suggest that patient memory for treatment is a multifaceted construct. The three questions included in the Conceptualization Task, prompting patients to identify contributing factors to psychopathology, make intervention recommendations, and provide a rationale for recommendations were not correlated enough with each other to be considered one unified scale. The Conceptualization Task showed the strongest associations with the Patient Recall Task, but none of the correlations with existing measures of patient memory for

treatment exceeded the convergent validity benchmark of $r \geq 0.50$ (Abma et al., 2016).

The Conceptualization Task may not show strong associations with the Patient Recall Task because the Conceptualization Task involves recalling *and applying* treatment contents to a hypothetical person, which represents a more advanced stage of learning relative to simply recalling treatment contents (Chi & Wylie, 2014; Lee & Harvey, 2015). The Conceptualization Task may show stronger associations with the Patient Recall Task than with the Generalization Task and Thoughts and Application Tasks because both the Conceptualization Task and Patient Recall Task are measures of memory for treatment contents in general, regardless of how much patients are recalling and implementing treatment contents in their daily lives. In contrast, the Generalization Task and Thoughts and Application Task assess how much patients are remembering and applying treatment contents *to themselves in their daily lives*, which represents a more advanced stage in the behavior change process relative to recalling treatment contents in general (Gumport et al., 2015, 2018; Michie et al., 2011).

Thus, the three questions in the Conceptualization Task may measure aspects of patient memory for treatment that are distinct from each other and from existing measures of patient memory for treatment. Relatedly, previous research also indicates that the Thoughts and Application Task and Generalization Task are relatively distinct from the Patient Recall Task, with correlations ranging from -0.04 to 0.48 (Gumport et al., 2018). This finding further supports the multifaceted nature of patient memory for treatment. However, it is important to note that correlations between these measures should be viewed as conservative tests of convergent validity because each measure uses a different format, and thus all common method variance is excluded (Podsakoff et al., 2003). Existing interpretation guidelines for convergent validity do not account for common method variance (Abma et al., 2016).

Findings from this study supported the discriminant validity of the Conceptualization Task, suggesting that performance on the task was distinct from, but may be influenced by, educational attainment, general working memory functioning, and general declarative memory functioning. These results could suggest, for example, that general memory functioning influences how much patients remember from treatment. Alternatively, general memory functioning may influence how patients engage with the logistics of the Conceptualization Task, which would constitute a source of systematic measurement error (e.g., patients with better general memory functioning are more able to recall and respond appropriately to the text vignette) (Brakenhoff et al., 2018). This study aimed to reduce demands on general memory functioning during the Conceptualization Task by allowing patients to re-read the vignettes while responding

to questions for the task. In any case, future research should clarify this possible source of measurement error in the Conceptualization Task.

Criterion validity analyses suggested that therapist use of constructive memory support strategies, but not non-constructive strategies, predicted performance on several aspects of the Conceptualization Task at post-treatment. These results reflect findings from the educational psychology literature favoring constructive learning activities over non-constructive activities for promoting memory and learning (Chi & Wylie, 2014; Menekse et al., 2013), and build on emerging evidence in the context of the MSI that constructive memory support strategies may be more effective at promoting patient memory for treatment than non-constructive memory support strategies (Zieve, et al., 2019b).

Notably, the relation between therapist use of constructive memory support strategies and patient performance on the Conceptualization Task did not remain statistically significant at the 6-month follow-up assessment point. This finding corroborates results from a previous study using the Patient Recall Task, which also found that the effect of constructive memory support on patient memory for treatment was only statistically significant at post-treatment and not follow-up time points (Zieve, et al., 2019b). These results suggest that the effect of the MSI may be more limited to time periods closer to treatment (e.g., during treatment and immediately after treatment). Other factors (e.g., patient adherence to treatment) may become more important influences on patient memory for treatment in the months and years after treatment has ended.

Contrary to hypotheses, one continuous memory support variable describing the average number of different types of memory support strategies therapists used per session showed a negative association with performance on the Conceptualization Task at the 6-month follow-up. This finding is counterintuitive given that strategies from the MSI derive from multiple, complementary theories of memory and are hypothesized to target distinct processes, suggesting that using a variety of strategies would result in a more comprehensive intervention. For example, the strategies of application and evaluation may enhance memory by encouraging patients to engage in deeper *levels of processing* (Craik & Lockhart, 1972), whereas the strategies of practice remembering and cue-based reminder may take effect by expanding *retrieval routes* to stored information (Bjork, 1988).

Several mechanisms may account for this finding. Attempting to engage multiple memory processes by using a variety of memory support strategies may not engage each memory process to a sufficient degree. Alternatively, attempting to use a greater variety of memory support strategies may result in a higher cognitive load for therapists (van Merriënboer & Sweller, 2005), potentially leading to less effective implementations of individual strategies. Given

the preliminary nature of these findings, additional studies are needed to replicate this effect and clarify the mechanism of action. One valuable direction for future research would be examining the relations between each individual memory support strategy in the MSI and key outcomes (e.g., patient memory for treatment, treatment outcome). This could clarify whether some memory support strategies lend themselves more to cognitive therapy treatment contents than others.

Further criterion validity analyses indicated that only performance on second question of the Conceptualization Task, which assessed patient ability to make treatment-consistent intervention recommendations, showed any relation to treatment outcome. However, it is important to note that these findings did not remain statistically significant after correcting for multiple comparisons. These results suggest that patient ability to make treatment-consistent intervention recommendations may be more important for treatment outcome than patient ability to identify contributing factors to psychopathology and provide a rationale for intervention recommendations. This extends previous findings that patient memory for *procedural* treatment contents is more closely related to treatment outcome than patient memory for *conceptual* treatment contents (Richardson et al., 2019). Whereas conceptual treatment contents describe general principles and theories supporting treatment recommendations (e.g., the cognitive model), procedural treatment contents describe the intervention recommendations themselves (e.g., cognitive restructuring, behavioral activation) (de Jong & Ferguson-Hessler, 1996). Future research could explore narrowing the focus of the Conceptualization Task to only memory for procedural treatment contents, for example, by retaining only question two. These results also raise the possibility that targeting MSI strategies on procedural treatment contents may result in better outcomes than spreading MSI strategies across procedural and conceptual treatment contents.

Notably, the relation between making treatment-consistent intervention recommendations on the Conceptualization Task and treatment outcome was no longer statistically significant after controlling for existing measures of patient memory for treatment, suggesting that the Conceptualization Task did not predict any independent variance in treatment outcome. Of the three questions included in the Conceptualization Task, the question that assessed patient ability to make treatment-consistent intervention recommendations was most similar to existing measures of patient memory for treatment. Indeed, patients often answered the intervention recommendation question from the Conceptualization Task in a comparable manner to the Patient Recall Task (e.g., by listing a number of specific skills and strategies). Given these results, the primary advantage of the Conceptualization Task over the Patient Recall Task is that the

Conceptualization Task can be administered at baseline to assess pre-treatment knowledge of treatment contents, whereas the framing of the Patient Recall Task requires that some sessions have occurred.

This study involved several additional limitations. First, the Conceptualization Task was collected at relatively few timepoints, limiting the sophistication of analyses that could be conducted. For example, if the task were collected at one additional timepoint in the middle of treatment, then parallel process growth curve modeling (Little, 2013) could be conducted to test whether the trajectory of change in Conceptualization Task performance is related to the trajectory of change in treatment outcome across treatment. Second, many of the analyses conducted for this study were correlational, and as such, causation cannot be inferred. Third, this study was limited to a sample of patients receiving cognitive therapy for depression. Broader application across a range of mental health conditions and interventions is in order. Fourth, although the Conceptualization Task was designed to measure patient ability to conceptualize psychological problems in a way that reflects accurate memory of the theoretical model used in treatment, we did not confirm through observational methods that all treatment providers discussed the theoretical model with patients. However, routinely discussing the theoretical model was part of the treatment protocol. Finally, results from this study were based on single items from the Conceptualization Task. It was not possible to assess internal consistency for each individual item, although other metrics of reliability (inter-rater and test–retest) were presented.

Conclusion

This study provided mixed preliminary evidence supporting the reliability and validity of the Conceptualization Task as a measure of patient memory for treatment. However, further refinement and testing is needed to meet psychometric standards. Future versions of the Conceptualization Task may contribute to assessing the multifaceted aspects of patient memory for treatment, particularly when assessment of baseline knowledge of treatment contents is needed.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10608-022-10293-1>.

Author Contributions GZ and AH contributed to the study conception and design. Material preparation, data collection and analysis were performed by GZ, CA, IR and SG. The first draft of the manuscript was written by GZ and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was funded by the National Institute of Mental Health (Grant Number R01MH108657).

Data Availability The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interest Garret G. Zieve, Courtney C. Armstrong, Ian M. Richardson, Sydney B. Garcia, Allison G. Harvey have no conflict of interest.

Ethical Approval The methodology for this study was approved by the Committee for the Protection of Human Subjects of the University of California, Berkeley (Ethics approval number: 2011-11-3795).

Informed Consent Informed consent was obtained from all individual participants included in the study.

Animal Rights No animal studies were carried out for this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abma, I. L., Rovers, M., & Van Der Wees, P. J. (2016). Appraising convergent validity of patient-reported outcome measures in systematic reviews: Constructing hypotheses and interpreting outcomes. *BMC Research Notes*, 9(1), 226–231. <https://doi.org/10.1186/s13104-016-2034-2>
- Andersson, G., Carlbring, P., & Furmark, T. (2012). Therapist experience and knowledge acquisition in internet-delivered CBT for social anxiety disorder: A randomized controlled trial. *PLoS ONE*, 7(5), e37411. <https://doi.org/10.1371/journal.pone.0037411>
- Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond*. Guilford Press.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1), 289–300.
- Berg, M., Rozentel, A., Johansson, S., Liljethörn, L., Radvugin, E., Topooco, N., & Andersson, G. (2019). The role of knowledge in internet-based cognitive behavioural therapy for adolescent depression: Results from a randomised controlled study. *Internet Interventions*, 15, 10–17. <https://doi.org/10.1016/j.invent.2018.10.001>
- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (pp. 396–401). Wiley.
- Brakenhoff, T. B., Mitroiu, M., Keogh, R. H., Moons, K. G. M., Groenwold, R. H. H., & Van Smeden, M. (2018). Measurement error is often neglected in medical literature: A systematic review. *Journal of Clinical Epidemiology*, 98, 89–97. <https://doi.org/10.1016/j.jclinepi.2018.02.023>
- Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: Meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, 26(4), 543–570. <https://doi.org/10.1080/13854046.2012.680913>
- Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Brookline Books.
- Chambers, M. J. (1991). Patient recall of recommendations in the behavioural treatment of insomnia. *Sleep Research*, 20, 222.
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49, 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290. <https://doi.org/10.1037//1040-3590.6.4.284>
- Cortina, J. M. (1993). What Is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.
- de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, 21, 105–113.
- Dong, L., Zhao, X., Ong, S. L., & Harvey, A. G. (2017). Patient recall of specific cognitive therapy contents predicts adherence and outcome in adults with major depressive disorder. *Behaviour Research and Therapy*, 97, 189–199. <https://doi.org/10.1016/j.brat.2017.08.006>
- Eells, T. D., Lombart, K. G., Kendjelic, E. M., Turner, L. C., & Lucas, C. P. (2005). The quality of psychotherapy case formulations: A comparison of expert, experienced, and novice cognitive-behavioral and psychodynamic therapists. *Journal of Consulting and Clinical Psychology*, 73(4), 579–589. <https://doi.org/10.1037/0022-006X.73.4.579>
- Ellis, T. D. (Ed.). (2007). *Handbook of psychotherapy case formulation* (2nd ed.). The Guilford Press.
- Enders, C. K. (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. *Behaviour Research and Therapy*, 98, 4–18. <https://doi.org/10.1016/j.brat.2016.11.008>
- First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). *Structured clinical interview for DSM-5-research version (SCID-5-RV)*. American Psychiatric Association.
- Greenberger, D., & Padesky, C. A. (2015). *Mind over mood: Change how you feel by changing the way you think*. The Guilford Press.
- Gumport, N. B., Williams, J. J., & Harvey, A. G. (2015). Learning cognitive behavior therapy. *Journal of Behavior Therapy and Experimental Psychiatry*, 48, 164–169. <https://doi.org/10.1016/j.jbtep.2015.03.015>
- Gumport, N. B., Dong, L., Lee, J. Y., & Harvey, A. G. (2018). Patient learning of treatment contents in cognitive therapy. *Journal of Behavior Therapy and Experimental Psychiatry*, 58, 51–59. <https://doi.org/10.1016/j.jbtep.2017.08.005>
- Hahlweg, K., & Richter, D. (2010). Prevention of marital instability and distress: Results of an 11-year longitudinal follow-up study. *Behaviour Research and Therapy*, 48(5), 377–383. <https://doi.org/10.1016/j.brat.2009.12.010>
- Harvey, A. G., Dong, L., Lee, J. Y., Gumport, N. B., Hollon, S. D., Rabe-Hesketh, S., Hein, K., Haman, K., McNamara, M. E., Weaver, C., Martinez, A., Notsu, H., Zieve, G., & Armstrong, C. C. (2017). Can integrating the Memory Support Intervention into cognitive therapy improve depression outcome? Study protocol for a randomized controlled trial. *Trials*, 18(1), 1–11. <https://doi.org/10.1186/s13063-017-2276-x>

- Harvey, A. G., Lee, J. Y., Williams, J. J., Hollon, S. D., Walker, M. P., Thompson, M. A., & Smith, R. (2014). Improving outcome of psychosocial treatments by enhancing memory and learning. *Perspectives on Psychological Science*, 9, 161–179. <https://doi.org/10.1177/1745691614521781>
- Harvey, A. G., Lee, J., Smith, R. L., Gumpert, N. B., Hollon, S. D., Rabe-Hesketh, S., Hein, K., Dolsen, M. R., Haman, K. L., Kanady, J. C., Thompson, M. A., & Abrons, D. (2016). Improving outcome for mental disorders by enhancing memory for treatment. *Behaviour Research and Therapy*, 81, 35–46. <https://doi.org/10.1016/j.brat.2016.03.007>
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55(4), 352–358.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kronmüller, K., Victor, D., Schenkenbach, C., Postelnicu, I., Backenstrass, M., Schröder, J., & Mundt, C. (2007). Knowledge about affective disorders and outcome of depression. *Journal of Affective Disorders*, 104, 155–160. <https://doi.org/10.1016/j.jad.2007.02.010>
- Lee, J. Y., & Harvey, A. G. (2015). Memory for therapy in bipolar disorder and comorbid insomnia. *Journal of Consulting and Clinical Psychology*, 83, 92–102. <https://doi.org/10.1037/a0037911>
- Lee, E. C., Whitehead, A. L., Jacques, R. M., & Julious, S. A. (2014). The statistical interpretation of pilot trials: Should significance thresholds be reconsidered? *BMC Medical Research Methodology*, 14, 41. <https://doi.org/10.1186/1471-2288-14-41>
- Lee, J. Y., Worrell, F. C., & Harvey, A. G. (2016). The development and validation of the memory support rating scale (MSRS). *Psychological Assessment*, 28, 715–725. <https://doi.org/10.1037/pas0000219>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. The Guilford Press.
- Mander, B. A., Santhanam, S., Saletin, J. M., & Walker, M. P. (2011). Wake deterioration and sleep restoration of human learning. *Current Biology*, 21(5), R183–R184. <https://doi.org/10.1016/j.cub.2011.01.019>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. H. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, 102, 346–374. <https://doi.org/10.1002/jee.20021>
- Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(1), 42. <https://doi.org/10.1186/1748-5908-6-42>
- Miller, S. L., Celone, K., DePeau, K., Diamond, E., Dickerson, B. C., Rentz, D., Pihlajamäki, M., & Sperling, R. A. (2008). Age-related memory impairment associated with loss of parietal deactivation but preserved hippocampal activation. *PNAS*, 105(6), 2181–2186. <https://doi.org/10.1073/pnas.0706818105>
- Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, 33(3), 484–499. <https://doi.org/10.1016/j.cpr.2013.01.010>
- Myles, P. J., & Milne, D. L. (2004). Outcome evaluation of a brief shared learning programme in cognitive behavioural therapy. *Behavioural and Cognitive Psychotherapy*, 32, 177–188. <https://doi.org/10.1017/S1352465804001183>
- Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology*, 15(6), 1044–1045. <https://doi.org/10.1093/beheco/arh107>
- Nelson, H. E., & Willison, J. (1991). *National adult reading test (NART)*. Nfer-Nelson.
- Persons, J. B. (2008). *The case formulation approach to cognitive-behavior therapy*. The Guilford Press.
- Podsakoff, P. M., Mackenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Richardson, I. M., Zieve, G. G., Harvey, A. G. (2019). *Conceptual and procedural treatment contents in CBT for procrastination: Differential client recall and effects of recall on outcome*. In Poster presentation at the meeting of the Association for Behavioral and Cognitive Therapies, Atlanta, GA.
- Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The inventory of depressive symptomatology (IDS): Psychometric properties. *Psychological Medicine*, 26, 477–486. <https://doi.org/10.1017/S0033291700035558>
- Rush, A. J., Kraemer, H. C., Sackeim, H. A., Fava, M., Trivedi, M. H., Frank, E., Ninan, P. T., Thase, M. E., Gelenberg, A. J., Kupfer, D. J., Regier, D. A., Rosenbaum, J. F., Ray, O., & Schatzberg, A. F. (2006). Report by the ACNP task force on response and remission in major depressive disorder. *Neuropsychopharmacology*, 31, 1841–1853. <https://doi.org/10.1038/sj.npp.1301131>
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 501–518. <https://doi.org/10.1037/0278-7393.13.3.501>
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117(1), 34–50.
- Sperling, R. A., Bates, J. F., Chua, E. F., Cocchiarella, A. J., Rentz, D. M., Rosen, B. R., Schacter, D. L., & Albert, M. (2003). fMRI studies of associative encoding in young and elderly controls and mild Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, 74(1), 44–50. <https://doi.org/10.1136/jnnp.74.1.44>
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford University Press.
- Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T., Crismon, M. L., Shores-Wilson, K., Toprac, M. G., Dennehy, E. B., Witte, B., & Kashner, T. M. (2004). The Inventory of depressive symptomatology, clinician rating (IDS-C) and self-report (IDS-SR), and the quick inventory of depressive symptomatology, clinician rating (QIDS-C) and self-report (QIDS-SR) in public sector patients with mood disorders: A psy. *Psychological Medicine*, 34(1), 73–82.
- Üstün, T. B., Chatterji, S., Kostanjsek, N., Rehm, J., Kennedy, C., Epping-Jordan, J., Saxena, S., Korff, M. V., & Pull, C. (2010). Developing the World Health Organization disability assessment schedule 2.0. *Bulletin of the World Health Organization*, 88, 815–823. <https://doi.org/10.2471/BLT.09.067231>
- van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2), 147–177. <https://doi.org/10.1007/s10648-005-3951-0>
- Young, J., & Beck, A. T. (1980). *Cognitive therapy scale: Rating manual*. Unpublished manuscript, University of Pennsylvania, Philadelphia.

- Zieve, G. G., Dong, L., & Harvey, A. G. (2019a). Patient memory for psychological treatment contents: Assessment, intervention, and future directions for a novel transdiagnostic mechanism of change. *Behaviour Change*, 36(1), 1–11. <https://doi.org/10.1017/bec.2019.1>
- Zieve, G. G., Dong, L., Weaver, C., Ong, S. L., & Harvey, A. G. (2019b). Patient constructive learning behavior in cognitive therapy: A pathway for improving patient memory for treatment? *Behaviour Research and Therapy*, 116, 80–89. <https://doi.org/10.1016/j.brat.2019.02.006>
- Zieve, G. G., Woodworth, C., & Harvey, A. G. (2020). Client memory and learning of treatment contents: An experimental study

of intervention strategies and relationship to outcome in a brief treatment for procrastination. *Journal of Behavior Therapy and Experimental Psychiatry*, 69, 101579. <https://doi.org/10.1016/j.jbtep.2020.101579>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.