Title
The Development and Validation of the Memory Support Rating Scale

Journal
Psychological Assessment, 28(6)

ISSN
1040-3590

Authors
Lee, Jason Y
Worrell, Frank C
Harvey, Allison G

Publication Date
2016-06-01

DOI
10.1037/pas0000219

Peer reviewed

HHS Public Access
Author manuscript
Psychol Assess. Author manuscript; available in PMC 2016 June 01.

Published in final edited form as:
Psychol Assess. 2016 June ; 28(6): 715–725. doi:10.1037/pas0000219.

The Development and Validation of the Memory Support Rating Scale (MSRS)

Jason Y. Lee, **Frank C. Worrell**, and **Allison G. Harvey**
University of California, Berkeley

Abstract

Patient memory for treatment information is poor, and worse memory for treatment information is associated with poorer clinical outcomes. Memory support techniques have been harnessed to improve patient memory for treatment. However, a measure of memory support used by treatment providers during sessions has yet to be established. The present study reports on the development and psychometric properties of the Memory Support Rating Scale (MSRS) – an observer-rated scale designed to measure memory support. Forty-two adults with major depressive disorder (MDD) were randomized to either cognitive therapy plus memory support (CS+MS; $n = 22$) or cognitive therapy as-usual (CT-as-usual; $n = 20$). At post-treatment, patients freely recalled treatment points via the Patient Recall Task. Sessions ($n = 171$) were coded for memory support using the MSRS, 65% of which were also assessed for the quality of cognitive therapy via the Cognitive Therapy Rating Scale (CTRS). A unidimensional scale composed of 8 items was developed using exploratory factor analysis, though a larger sample is needed to further assess the factor structure of MSRS scores. High inter-rater and test-retest reliabilities of MSRS scores were observed across seven MSRS coders. MSRS scores were higher in the CT+MS condition compared to CT-as-usual, demonstrating group differentiation ability. MSRS scores were positively associated with Patient Recall Task scores but not associated with CTRS scores, demonstrating convergent and discriminant validity, respectively. Results indicate that the MSRS yields reliable and valid scores for measuring treatment providers' use of memory support while delivering cognitive therapy.

Keywords

memory support; therapist technique; scale development; cognitive therapy; depression

---

Memory impairment is prevalent across a wide range of mental illnesses. Memory deficits have been documented in major depression (e.g., Behnken et al., 2010), bipolar disorder (e.g., Martino, Igoa, Marengo, Scarpola, & Strejilevich, 2011), schizophrenia (e.g., Varga, Magnusson, Flekkoy, David, & Opjordsmoen, 2007), post-traumatic stress disorder (e.g., Isaac, Cushway, & Jones, 2006), and the anxiety disorders (e.g., Airaksinen, Larsson, & Forsell, 2005). These deficits have been attributed to a variety of difficulties, including the organization of information during learning (Behnken, et al., 2010), problems spontaneously

---

Corresponding Author: Allison G. Harvey, University of California, Berkeley, Department of Psychology, 2205 Tolman Hall #1650, Berkeley, CA 94720-1650. Phone: (510) 642-7138. ; Email: aharvey@berkeley.edu

using memory-promoting strategies, reduced hippocampal volume (e.g., Videbech & Ravnkilde, 2004), narrowing of attention and cognitive capacity associated with negative emotion (Gotlib & Joormann, 2010), and cognitive impairment associated with comorbid sleep impairment (Yoo, Hu, Gujar, Jolesz, & Walker, 2007).

Perhaps not surprisingly, patient memory for the content of treatment information is strikingly poor. Chambers (1991) reported that patients with insomnia forget one third of the instructions given during behavioral therapy for insomnia; for some types of recommendations, recall was as low as 13%. Moreover, Lee and Harvey (2015) found that patients diagnosed with bipolar disorder and co-morbid insomnia recalled only between 19.6% and 36.9% of treatment points listed by treatment providers from one weekly session to the next. These findings are consistent with related literature examining patient memory for medical advice, which feature recall rates between 19% and 33% within the months following a clinic visit (e.g., Bober, Hoke, Duda, & Tung, 2007; Jansen et al., 2008; Lewkovich & Haneline, 2005; Pickney & Arnason, 2005).

There is emerging evidence that poor memory for treatment is associated with worse treatment outcome. Lee and Harvey (2015) found a positive correlation between patient recall and sleep outcome following cognitive behavior therapy or psychoeducation treatments for insomnia among patients diagnosed with bipolar disorder. Similarly, poor memory for the content of a doctor's visit leads to lower treatment adherence (e.g., Bober et al., 2007; Jansen et al., 2008; Tosteson et al., 2003), which in turn leads to the incorrect or incomplete compliance with treatment recommendations (Ley et al., 1976), and subsequently worse outcome (e.g., Simpson et al., 2011).

Given that memory for treatment is poor, and failing to remember treatment information is associated with worse treatment outcomes, Harvey et al. (2014) have hypothesized that patients would benefit from the use of memory support techniques utilized by treatment providers during treatment sessions. Indeed, the accumulating evidence suggests that memory support techniques can improve memory encoding and retention in a wide range of populations, including patients with Alzheimer's disease and vascular dementia (Almkvist, Fratiglioni, Agüero-Torres, Viitanen, & Bäckman, 2010), older adults and individuals with poor frontal lobe function (Bunce, 2003), as well as patients with depression (Bäckman & Forsell, 1994; Taconnat et al., 2010). Also, based on a thorough review of the cognitive psychology and education literatures and carefully crafted criteria, specific memory support strategies for use during treatment have been identified (Harvey et al., 2014).

A significant obstacle to further progress in this area of research is the absence of a validated measure to systematically assess the use of memory support by treatment providers. The development of such a measure would provide a means to monitor the use of memory support in treatment, establish the dose of memory support needed to optimize outcomes, and examine if and how changing levels of memory support might play a causal role in treatment outcome.

The overall goal of the current study is to develop an observer-reported measure of memory support – the Memory Support Rating Scale (MSRS) – and assess the psychometric properties of MSRS scores. The MSRS is intended to be transdiagnostic, with the goal of being useful across a wide range of treatments and psychiatric disorders. In the present study, however, the focus was on a sample of individuals receiving cognitive therapy for depression as a platform for developing the MSRS. The rationale for focusing on depression is twofold. First, depression is highly prevalent and impairing, and is thus an important mental health concern (Moussavi et al., 2007). Second, deficits in declarative and working memory – which may be particularly important for the ability to recall therapeutic information – are common in depression (for a review, see Snyder, 2013). The rationale for focusing on cognitive therapy is based on the substantial and promising evidence for the efficacy of cognitive therapy for depression (DeRubeis et al., 2005), yet there is also room for improvement in outcome (Dimidjian et al., 2006).

This study has four specific aims. The first is to develop the MSRS and assess the internal consistency and structural validity of MSRS scores. The second is to establish the inter-rater reliability and test-retest reliability of MSRS scores. The third is to establish the group differentiation ability of MSRS scores by examining whether these scores are able to detect experimental manipulations of memory support. The hypothesis tested is that higher MSRS scores will be observed among participants receiving cognitive therapy enhanced with memory support (CT+MS) compared to participants receiving cognitive therapy as usual (CT-as-usual). The fourth is to further establish the construct validity of MSRS scores by assessing the relation between MSRS scores and patient recall for treatment contents, measured by the Patient Recall Task, as well as observer-rated evaluations of quality of cognitive therapy, measured by the Cognitive Therapy Rating Scale (CTRS). Given that memory support targets patient memory for treatment information, and is not intended to influence the fundamental aspects of cognitive therapy itself, MSRS scores were expected to be positively associated with Patient Recall Task scores (convergent validity) but unrelated to CTRS scores (discriminant validity).

Method

Development of the MSRS

The MSRS was devised via a careful iterative process. First, strategies demonstrated to be effective in improving the encoding, storage, and retrieval stages of a memory were identified via a thorough review of the cognitive and education literatures (Harvey et al., 2014). Second, tapes of sessions in which cognitive therapy for depression was delivered were coded to determine which of the identified strategies are utilized, or have the potential to be utilized, by treatment providers. Third, similar strategies were consolidated in order to reduce redundancy. This process resulted in the following eight non-overlapping memory support strategies.

Attention recruitment—This strategy involves the treatment provider’s use of expressive language that explicitly communicates to the patient that a treatment point is important to remember (e.g., “if there is one thing to remember in ten years time, it is this skill”; “this is a key point to remember”), or multimedia/diverse presentation modes (e.g., handouts, poems, songs, note taking, role-playing, imagery, using a white board) as a means to recruit and engage the patient's attention. Strategies that promote attention recruitment have been demonstrated to be particularly important for the encoding process of a memory (e.g., Carney & Levin, 2002; Craik, Govini, Naveh-Benjamin, & Anderson, 1996).

Categorization—This strategy involves explicit effort by the treatment provider to work with the patient to categorize 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."). There is strong evidence that organizing information into chunks can be effective in improving memory (e.g., Taconnat et al., 2010).

Evaluation—This strategy involves the treatment provider working with the patient to (a) discuss the pros and cons of a treatment point (e.g., "What are some advantages and 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 all day when you are feeling depressed?"). Evaluating information has been demonstrated to be effective in promoting the encoding and retention of information (e.g., Williams & Lombrozo, 2010; Williams, Lombrozo, & Rehder, 2013).

Application—This strategy involves the treatment provider 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?"). Ample empirical demonstrations indicate that taking principles learned in one situation and applying them to another situation improves learning (e.g., Kolodner, 1997; Mestre, 2005).

Repetition—This strategy involves the treatment provider restating, rephrasing, or revisiting information discussed in treatment (e.g., "in other words," "as we talked about earlier," or "in sum"). There is robust evidence that repetition helps automatize new knowledge and that temporally spaced repetition of information is effective in the learning process (e.g., Siegel, & Kahana, 2014).

Practice remembering—This strategy involves the treatment provider facilitating the patient to regenerate, restate, rephrase, and/or revisit a treatment point (e.g., "Can you tell me what some of the main ideas you've taken away from today's session?). Studies have documented the effectiveness of testing individuals on learned material as a means to promote further learning and memory of the same material (e.g., Karpicke & Roediger, 2008). Moreover, learning is clearly improved if practice remembering is spread out in time rather than massed together (e.g., Kornell, Castel, Eich, & Bjork, 2010).

Cue-based reminder—This strategy involves the treatment provider helping the patient develop new or existing cues (e.g., colored wrist bands, reminder text messages/phone calls/e-mails, smart phone apps, acronyms, rhymes, and other mnemonics) to facilitate memory for treatment points. Strategies that involve cue-based reminders have been linked to improved memory for target information (e.g., Gollwitzer & Sheeran, 2006; Kapur, Glisky, & Wilson, 2004).

Praise recall—This strategy involves the treatment provider 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."). Studies have demonstrated the effectiveness of social praise in reinforcing learning (e.g., Gielen, Peeters, Dochy, Onghena, & Struyven, 2010).

Finally, definitions of memory support strategies were honed during the coding of hour-long treatment tapes by three independent raters. Codes were compared for each tape, and proposed revisions were discussed and established by consensus. Revisions to the scale were necessary for the first 15 tapes coded. Revisions to the scale were unnecessary for the following three tapes, at which point the assessment for psychometric properties began. See Appendix A for the MSRS.

Participants

Participants were 48 adults who met diagnostic criteria for major depressive disorder and were recruited to participate in an NIMH-funded randomized control trial of cognitive therapy for depression. All participants were recruited from the greater Alameda County to reflect population demographics. The University of California, Berkeley, Committee for the Protection of Human Subjects approved the study. All participants provided written informed consent and were financially compensated.

Participants were assessed via an in-person interview to meet the following inclusion criteria: (a) diagnosis of major depressive disorder (MDD), first episode, recurrent or chronic, according to DSM-IV-TR criteria (American Psychiatric Association, 2000); (b) a minimum score of 24 or above on the Inventory of Depressive Symptomatology - Self Report (IDS-SR; Rush, Gullion, Basco, Jarrett, & Trivedi, 1996); (c) a minimum age of 18 years; (d) a stable regimen of medications taken for mood (if any) for the past eight weeks, with such medications having a minimal effect on memory; and (e) an ability and willingness to give informed consent.

Participants were excluded for (a) a history of bipolar affective disorder; (b) a history of psychosis (including schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder, or psychotic organic brain syndrome); (c) a current non-psychotic Axis I disorder if it constitutes the principal diagnosis and if it requires treatment other than that offered in the project; (d) a history of substance dependence in the past six months; (e) antisocial, borderline, or schizotypal personality disorder; (f) an IQ score below 80; (g) evidence of any medical disorder or condition that could cause depression or preclude participation in cognitive therapy; and (h) current suicide risk sufficient to preclude treatment on an outpatient basis.

Procedure

Eligible participants were randomly allocated to receive one of two interventions: cognitive therapy plus memory support (CT+MS) or cognitive therapy as usual (CT-as-usual). Both treatment groups received weekly individual treatment sessions lasting 60 minutes for 14 consecutive weeks. Both conditions were matched for the number and quality of handouts. All treatment sessions were videotaped. Treatment providers in both conditions had a Master's or doctoral degree in psychology. The two conditions followed an identical protocol, with the exception that treatment providers in the CT+MS condition received training in the use of eight memory support strategies as identified above, and were directed to use as much memory support as possible while delivering treatment.

Twenty-five participants were randomly assigned to the CT+MS condition, and 23 participants were assigned to the CT-as-usual condition. A minimum of three randomly selected tapes per participant (including one randomly selected tape from each tertile of the temporal ordering of the 14 total sessions) were assigned to be coded for memory support by fully trained independent raters. Memory support raters were provided with the following coding instructions: *As you listen to a recorded session, pay close attention to the content of the dialogue as well as the behaviors of the patient and treatment provider. Keep notes to help you keep track of "treatment points" (i.e., insights, skills, or strategies that the treatment provider would want the patient to remember as part of the treatment). Make sure to assign a code only when the treatment provider is supporting memory for one or more treatment points. Also keep track of the timestamps at which memory support is utilized. Timestamps are to be recorded as soon as it is clear that a specific memory support strategy is utilized. For instance, the patient might be writing on the board a long time, but you should mark the time as soon as the therapist asks the patient to write on the board. To maximize coding accuracy, rewind, replay, and check for memory support as often as need be. While there is no limit on how many times a tape can be replayed while coding, a general rule of thumb is to spend about 120 minutes coding each 60-minute session.*

Memory support raters coded tapes for memory support as soon as they became available throughout the treatment trial. Memory support in a given session was scored in three ways: (a) the total number of instances in which memory support was utilized ("MS Instances"); (b) the amount of memory support utilized per minute, calculated by dividing MS Instances by the total length of the session in minutes ("MS Instances per Minute"); and (c) the total number of distinct memory support items utilized ("MS Items Used").

The quality of cognitive therapy was rated in a subsample of 75 tapes using the Cognitive Therapy Rating Scale (CTRS; Young & Beck, 1980). The CTRS has acceptable psychometric properties (Vallis, Shaw, & Dobson, 1986) and was rated by one of three clinical psychologists, all of who have training and experience in using the CTRS.

At post-treatment, participants completed the Patient Recall Task – a free recall task established in previous studies (Lee & Harvey, 2015). In this task, participants are given 10 minutes to write down as many treatment points as they could remember from the start of treatment up to (and including) their most recent session. Patient recall for a given session is then measured based on the raw number of treatment points freely recalled from the start of treatment to the end of Session 14.

Throughout the treatment trial, a subsample of forty-three 10-minute treatment tape segments was randomly selected for establishing inter-rater reliability. Pairs of memory support raters were randomly selected from the pool of seven raters (the six trained raters plus JL to independently code the subsample of tapes for memory support. To establish test-retest reliability, a subsample of thirty-two 10-minute treatment tape segments was randomly selected throughout the treatment trial. Memory support raters were randomly selected to independently code the subsample of tapes. Three months later, the same raters recoded the same subsample of tapes.

Training of Memory Support Raters

Six raters were trained to code the eight memory support techniques. Training was conducted using a multi-step approach. First, the team of raters was introduced to the MSRS and coded the same 60-minute treatment tape as a group along with JL. Second, the team of raters coded another 60-minute treatment tape as a group while JL coded the same tape independently. Third, raters were split into two groups of three, and each group coded another 60-minute treatment tape while JL coded the same tape independently. Fourth, raters were split into pairs, and each pair coded another 60-minute treatment tape while JL coded the same tape independently.

Throughout this training process, codes were compared and disagreements were discussed until consensus was achieved before proceeding to the next step. As a final step in the training, each rater was required to individually establish 80% or higher inter-rater agreement with JL across five consecutive 30-minute segments of treatment tapes. In line with similar coding systems (e.g., Specific Affect Coding System; Coan & Gottman, 2007), each minute of treatment observed was considered to be an opportunity in which raters can agree or disagree on the presence or absence of memory support. As such, agreement scores were calculated by dividing the number of minutes in which raters agree (defined as both raters having the same memory support code, or both raters having no memory support code) by the total length of the treatment tape (coded in minutes), and multiplying by 100%.

Measures

Inventory of Depressive Symptomatology - Self Report (IDS-SR)—The IDS-SR is a 30-item, clinician-rated measure of current depressive symptoms with total scores ranging from 0 to 84, and higher scores indicating greater depressive severity. IDS-SR scores have been demonstrated to have adequate reliability and validity (Rush et al., 1996).

The Structured Clinical Interview for DSM-IV-TR (SCID)—The SCID is a semi-structured interview designed to assess DSM-IV-TR diagnostic criteria for Axis I disorders (First, Spitzer, Gibbon, & Williams, 1995). SCID scores have been shown to have adequate reliability (Lobbestael, Leurgans, & Arntz, 2011). Trained psychology doctoral students administered the SCID to all participants to assess current and lifetime Axis I disorders and to confirm a diagnosis of major depressive disorder. All assessment sessions were tape recorded, and a randomly subset of tapes (20%) were selected for close scrutiny by independent reviewers blind to treatment condition and diagnoses. Inter-rater reliability for the primary diagnosis of major depressive disorder was good ($\kappa = 0.63$) according to Altman's (1991) guidelines.

Cognitive Therapy Rating Scale (CTRS)—The CTRS (Young & Beck, 1980) is an observer-rated scale measuring therapist competence in the use of cognitive therapy for depression. The CTRS contains 11 items rated on a Likert scale, with total scores ranging from 0 to 66, and higher scores indicating greater quality of cognitive therapy. CTRS scores have been demonstrated to accurately discriminate between acceptable and unacceptable therapist performance ratings made by independent observers, thereby exhibiting adequate discriminant validity (Vallis et al., 1986). In addition, acceptable inter-rater reliability has been established for CTRS scores (e.g., Vallis et al., 1986; Williams, Moorey & Cobb, 1991). In the present sample, the inter-rater reliability among random pairs of coders across $n = 14$ (18.7%) of the sessions coded for CTRS was ICC(1,1) = .77, indicating fair inter-rater reliability according to Cicchetti's (1994) guidelines.

Patient Recall Task—Careful consideration was given to devising this task. A recognition task was not selected to exclude the possibility of serving as a memory prompt, as demonstrated by Bäckman and Forsell (1994). Hence, a free recall measure based on previous studies (Lee & Harvey, 2015) was designed for the present study. As reported by Lee and Harvey (2015), the Patient Recall Task demonstrated adequate convergent validity (r's = .26 – .69) and strong inter-rater reliability ($r = .82$). This task was modified for the present study to assess for free recall of treatment contents across the entire treatment period, rather than assessing for recall of treatment contents from the past session only. The rationale for this change was to capture recall across the treatment period while limiting the possibility of weekly recall tasks inadvertently serving as a form of memory support.

The newly developed recall task consisted of a sheet of paper, which started with the following instructions: *Take a moment to think back to all the treatment sessions you've had with us so far. In the space provided below (use back of sheet if needed), please list as many distinct 'treatment points' as you can recall since the start of your treatment. A 'treatment point' is an insight, skill, or strategy that you think is important for you to remember and/or implement as part of your treatment. Make sure to only include points that are broad in scope (i.e., points that you would want to remember years from now). You have 10 minutes for this task. Please take the entire 10 minutes so that you record every single point you remember.*

Data Analysis Plan

Baseline differences in participant characteristics between treatment conditions were assessed via χ^2 and t-tests. Internal consistency of MSRS scores was assessed by examining inter-item correlations of MSRS scores (instances of each item) and Cronbach's α using the sample of tapes coded for memory support. Factor structure of the MSRS was established using exploratory factor analysis (EFA) across the sample of tapes coded for memory support. The suitability of the data for factor analysis was measured via two indices: the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1974) and Bartlett's test of sphericity (Bartlett, 1950). Common factor analysis (principal axis extraction) was used to explore latent factors in the set of 8 MSRS item scores. Common factor analysis was selected over principal components analysis as it has the advantage of accounting for measurement error in producing a factor solution (Widaman, 1993).

Multiple criteria were used to determine the number of factors to retain based on the recommendations of Henson and Roberts (2006). First, the sample of MSRS scores was assessed using Kaiser's (1960) criterion, which involves retaining factors with observed eigenvalues greater than 1. Second, the sample of MSRS scores was examined using Cattell's (1966) scree test, which involves evaluating a visual representation of eigenvalues ordered from largest to smallest and retaining only those factors that fall above a steep drop off point. Third, the sample of MSRS scores was evaluated using Horn's (1965) parallel analysis, which involves retaining factors with observed eigenvalues exceeding specified parameters of expected eigenvalues (based on a simulated distribution of eigenvalues from randomly generated samples). Horn's parallel analysis was conducted using the syntax developed by O'Conner (2000) in SPSS Version 22. Based on the recommendations of Glorfeld (1995), factors with observed eigenvalues that exceeded the 95th percentile of randomly generated eigenvalues were retained.

Inter-rater reliabilities of MSRS scores were assessed across randomly selected pairs of MSRS raters (drawn from a pool of four fully-trained raters) coding a total of forty-three 10-minute tape segments. Ratings were made independently, and inter-rater agreement was assessed via one-way random, absolute agreement, single measures intra-class correlation coefficients (ICC's; Shrout & Fleiss, 1979). Test-retest reliabilities of MSRS scores were examined across thirty-two 10-minute tape segments. Each tape segment was coded by a randomly-selected rater (from a pool of fully-trained MSRS raters) at two time points spaced at least two months apart.

Group differentiation ability of MSRS scores was assessed by comparing MSRS scores (i.e., MS Instances, MS Items Used, and MS Instances per Minute) across treatment conditions via *t*-tests and an effect size indicator (Cohen's *d*). Convergent validity was assessed via correlations between MSRS scores and Patient Recall Task scores, and by examining differences in means of Patient Recall Task scores by treatment condition via *t*-tests and Cohen's *d* effect sizes. Discriminant validity was assessed via correlations between MSRS scores and CTRS scores, and by comparing means of CTRS scores by treatment condition via *t*-tests and Cohen's *d* effect sizes.

Results

Preliminary Data Analyses

Three participants dropped out from each condition throughout the treatment (12.0% of participants in CT+MS vs. 13.0% of participants in CT-as-usual; $\chi^2(1) = 0.01$, $p = .913$). The remaining 42 participants completed treatment. Characteristics of participants who completed the randomized control trial by treatment condition are presented in Table 1. No baseline differences were observed between groups. Similarly, no baseline differences were observed between groups among the full sample of 48 participants who were randomized to receive treatment.

At least three tapes (one randomly-selected tape from each tertile of the 14 sessions) were coded per participant, except in the case of one participant who completed treatment (two out of three randomly selected tapes coded due to technical error with one tape), and the six participants who dropped out of the study during the treatment phase (tapes coded before dropout were retained). In addition to the pre-selected tapes, 35 randomly selected tapes in MS+CT and 6 randomly selected tapes in CT-as-usual were coded. The rationale for this additional coding was to aid the ongoing training of therapists in the MS+CT condition, and for demonstration purposes while training new Memory Support raters. The combination of preselected and additional tapes coded for memory support resulted in the final sample of 171 tapes (106 in MS+CT and 65 in CT-as-usual). The additional tapes were included in the final sample of tapes to increase power for factor analysis. Across the 171 tapes, mean session length was 62.33 minutes (*SD* = 13.34). Mean MS Instances was 14.61 (*SD* = 11.50), mean MS Items Used was 4.29 (1.68), and mean MS Instances per Minute was 0.24 (*SD* = 0.18).

Internal Consistency and Factor Structure of MSRS Scores

Internal consistency—Inter-item correlations of MSRS scores (instances of each item) across the total sample of 171 tapes coded across treatment conditions are presented in Table 2. Mean inter-item correlations among MSRS scores were in the medium range ($r = .33$) based on Cohen's (1988) guidelines, and the internal consistency estimate for MSRS scores based on the eight MSRS items was acceptable (Cronbach's α = .77) according to Nunnally's (1978) guidelines.

Exploratory factor analysis—The sample of 171 tapes coded for memory support was used for EFA. The KMO index was .78, indicating "middling" sampling adequacy according to Kaiser's (1974) guidelines, and Bartlett's test of sphericity was significant, $\chi^2(28) = 412.81$, $p < .001$, indicating that the correlation matrix is factorable based on Bartlett's (1950) guidelines. Initial communalities for Application, Praise Recall, Practice Remembering, Cue-Based Reminder, Repetition, Attention Recruitment, Categorization, and Evaluation were in the low to moderate range: .46, .38, .33, .21, .60, .55, .14, and .36, respectively.

Using Kaiser's (1960) criterion, two factors had eigenvalues greater than one, indicating a two-factor solution. Based on Cattell's (1966) scree test, one factor had an eigenvalue above a visually determined drop-off point, indicating a one-factor solution. According to Horn's (1965) parallel analysis, one factor had an observed eigenvalue exceeding its corresponding expected eigenvalue, again indicating a single-factor solution. Given the evidence that Kaiser's criterion has a tendency to overestimate the number of factors (Zwick & Velicer, 1986), and because parallel analysis has been demonstrated to be more accurate than Kaiser's criterion and other commonly used criteria (for a review, see Ledesma & Valero-Mora, 2007), a one-factor solution was extracted.

Based on the recommendations of Tabachnick and Fidell (2001), coefficients of items on factors should be 0.32 or higher. Factor loadings were above this threshold for all MSRS items (Application = .70, Praise Recall = .63, Practice Remembering = .55, Cue-Based Reminder = .34, Repetition = .80, Attention Recruitment = .76, Categorization = .35, and Evaluation = .56). Extracted communalities for Application, Praise Recall, Practice Remembering, Cue-Based Reminder, Repetition, Attention Recruitment, Categorization, and

Evaluation were .53, .49, .33, .32, .69, .60, .13, and .43, respectively. The one-factor solution explained 43.2% of the variance.

Inter-Rater and Test-Retest Reliability of MSRS Scores

Inter-rater reliability—Inter-rater reliabilities were "fair" for all MSRS scores – MS Instances ICC(1,1) = .74; MS Items Used ICC(1,1) = .73; and MS Instances per Minute ICC(1,1) = .74) – based on Cicchetti's (1994) guidelines.

Test-retest reliability—Mean number of days between Time 1 and Time 2 was 134.25 (SD = 98.84). MSRS scores at Time 1 were significantly and positively correlated with MSRS scores at Time 2 – MS Instances, $r = .72$; $p < .001$; MS Items Used, $r = .69$, $p < .001$; and MS Instances per Minute, $r = .72$, $p < .001$) – indicating strong temporal stability of MSRS scores.

Other Construct Validity Evidence Supporting MSRS Scores

Group differentiation ability—MSRS scores by treatment condition are presented in Table 3. Significantly higher MSRS scores (i.e., MS Instances, MS Items Used, and MS Instances per Minute) were observed in the CT+MS condition compared to the CT-as-usual condition. All corresponding effect sizes were large based on Cohen's (1988) guidelines.

Convergent validity—Correlations between MSRS scores and Patient Recall Task scores are presented in Table 4. MS Items Used and MS Instances per Minute were significantly, positively, and meaningfully associated with Patient Recall Task scores. MS Instances were not significantly correlated with Patient Recall Task scores, even though a medium effect size in the expected direction was observed according to Cohen's (1988) guidelines. Means of Patient Recall Task scores by treatment condition are presented in Table 3. Participants in the CT+MS group recalled, on average, more treatment points at post-treatment compared to participants in the CT-as-usual group. This difference did not reach significance, but it approached a medium effect size in the expected direction.

Discriminant validity—Correlations between MSRS scores and CTRS scores are presented in Table 4. None of MSRS scores (MS Instances, MS Items Used, and MS Rate) were significantly or meaningfully correlated with CTRS scores. Means of CTRS scores by treatment condition are presented in Table 3. There were no significant or meaningful group differences in CTRS scores between treatment conditions.

Discussion

The overarching goal of this study was to develop and assess the psychometric properties of scores on the MSRS – a scale developed to measure the extent to which treatment providers utilize memory support during treatment sessions. The first aim of the present study was to design the MSRS and assess the factor structure of MSRS scores. An observer-rated scale comprised of eight empirically derived MSRS items was formed and used to score a sample of cognitive therapy for depression sessions. The results of EFA indicated a single factor solution, with the set of MSRS items collectively explaining 43.2% of the total variance. The eight MSRS items were highly inter-correlated, and the one-factor solution demonstrated acceptable internal consistency. Although there was no a priori hypothesis regarding potential latent factors (hence the decision to run an EFA instead of a confirmatory factor analysis), the resultant unidimensional structure is perhaps not surprising given that each memory support strategy was devised to promote patient memory for treatment information in a unique manner.

The second aim of the present study was to establish the inter-rater and test-retest reliabilities of MSRS scores. Inter-rater reliability is the level of agreement among raters independently scoring an event. Given a common scale and standardized coding procedures, inter-rater reliability can serve as a reflection of the extent to which a scale can be used consistently across multiple raters (DeVellis, 2012). Similarly, test-retest reliability is the level of agreement between ratings made by the same rater across two different time points. Assuming that raters are using the same scale and are following standardized coding procedures at both time points, and that sufficient time has passed to minimize practice effects, high test-retest reliability suggests that the ratings of a scale are temporally stable (DeVellis, 2012). Across subsamples of tape segments in the present study, acceptable levels of inter-rater reliability and test-retest reliability were observed for all three MSRS scores (i.e., MS Instances, MS Items Used, and MS Instances per Minute). Together, these findings provide preliminary evidence that the MSRS can be reliably used across trained raters, and consistently within trained raters.

The third aim was to begin establishing evidence of construct validity of MSRS scores beyond structural validity and internal consistency by assessing their group differentiation ability. Construct validity is an assessment of the extent to which a scale measures what it purports to measure (Cronbach & Meehl, 1955). One way of assessing for construct validity is to assess for differences in a scale's scores between groups that are expected to differ in the construct of interest. A scale with scores that reflect differences in an experimentally manipulated construct (in this case, memory support) between groups provides construct validity evidence for those scores (Benson, 1998; Bornstein, 2011). In the CT+MS condition, treatment providers were trained in the strategic use of memory support techniques, and were encouraged to use as much memory support as possible in sessions. In contrast, in the CT-as-usual condition, treatment providers delivered standard CT. Theoretically, the experimental manipulation of memory support should lead to different levels of memory support observed between groups. Consistent with this hypothesis, the results indicate that the MSRS detected differences in memory support between treatment conditions. Specifically, participants in the CT+MS condition received significantly higher amounts of memory support, a greater diversity of memory support strategies, and a higher rate of memory support per minute, with large effect sizes observed between groups for all three MS scores.

The fourth and final aim was to further establish the construct validity of MSRS scores by assessing their convergent and discriminant validity. Convergent validity is an aspect of construct validity, and is defined as the extent to which a measure is associated with the outcome of a theoretically related construct (Cronbach & Meehl, 1955). In the present study, greater MS scores were expected to be associated with a higher number of treatment points freely recalled by patients at post-treatment, as memory support involves strategies that treatment providers use to promote patients' memory for treatment points. As hypothesized, a greater diversity of memory support strategies used, and a higher rate of memory support instances per minute, was each correlated with higher patient recall. Although the amount of memory support was not significantly correlated with patient recall, a medium effect size in the hypothesized direction was observed. Additionally, participants in the CT+MS group, who received a higher amount of memory support, a greater diversity of memory support strategies, and a higher rate of memory support per minute, recalled 24.1% more treatment points compared to participants in the CT-as-usual group. Although this difference was not significant, it approached a medium effect size in the expected direction. These findings underscore the effectiveness of the memory support strategies in improving the encoding, storage, and retrieval stages of a memory (for a review, see Harvey et al., 2014). It is noteworthy that patient memory is improved with not just the amount of memory support but also the use of a diversity of types of memory support. The varied use of memory support might serve a role in helping the patient to process the same treatment information in multiple ways, thereby promoting a deeper understanding and richer encoding of the treatment material. For instance, repeating a treatment point three times for a patient (Repetition) is not likely to be as effective in improving patient recall compared to repeating it once (Repetition), applying it to a hypothetical situation (Application), and then discussing its advantages and disadvantages (Evaluation). Indeed, previous studies have demonstrated that the use of multiple strategies to process information leads to better recall compared to the use of singular strategies (e.g., Clark & Paivio, 1991), perhaps as a function of the greater levels of processing that the former involves (Craik & Lockahrt, 1972).

Discriminant validity is another aspect of construct validity, and is defined as the extent to which a scale is uncorrelated with a theoretically different construct (Campbell & Fiske, 1959). In the present study, MSRS scores were expected to be uncorrelated with the quality of cognitive therapy, assessed via the Cognitive Therapy Rating Scale. The rationale is that memory support involves the use of specific strategies aimed at promoting patient memory for the contents of treatment, and are not intended to interfere with the traditional components of cognitive therapy (e.g., setting an agenda, maintaining interpersonal effectiveness, and eliciting patient feedback). As hypothesized, correlations between each of the MSRS scores (MS Instances, MS Items Used, and MS Rate) and CTRS scores were not significant. Furthermore, there were no differences in CTRS scores between treatment conditions, indicating that experimental manipulation of memory support did not result in a significant difference in cognitive therapy quality ratings. Together, these findings indicate that MSRS scores "behave" in predictable ways (i.e., are positively associated with related constructs and not associated with unrelated constructs), thereby exhibiting further evidence of construct validity.

Limitations and Conclusion

Several important considerations should be taken when interpreting the results of the present study. First, the psychometric properties of the MSRS were tested using a sample of individuals receiving cognitive therapy for major depressive disorder. Future work is needed to assess whether the MSRS yields valid and reliable scores for a range of patient populations and theoretical orientations in order to increase the generalizability of the current study findings. Second, a relatively small sample of tapes was used to establish the factor structure of the MSRS. Although a unidimensional model was retained based on the converging findings of EFA, strong internal consistency across all eight MSRS item scores, and theoretical underpinnings, future studies should conduct confirmatory factor analyses on larger samples of tapes to further assess the fit of a one-factor solution.

Third, the patient recall task used for validity assessment in the present study was a modified version of a previously developed task. Although the scores of the original recall task have been demonstrated to have adequate convergent validity and high inter-rater reliability, the psychometric properties of the modified recall task have yet to be established. Fourth, memory for treatment information was measured using a free recall measure of treatment information, which might be a limited (and perhaps overly conservative) measure of patient memory for treatment points. Indeed, classic experiments have demonstrated that free recall tasks result in less information recalled compared to recognition and cued recall tasks (e.g., Hart, 1967; Tulving & Pearlstone, 1966). Future studies are needed to measure patient memory for treatment contents via alternative recall tasks (e.g., a task that assesses patients' ability to apply principles learned in treatment in response to a hypothetical scenario) to serve as additional forms of predictive validity for the MSRS.

Limitations notwithstanding, the results of the current study suggest that the MSRS is a psychometrically sound tool in measuring the extent to which treatment providers utilize memory support during treatment. This observer-rated tool provides a means to monitor treatment providers' use of memory support in treatment, establish the dose of memory support needed to optimize treatment outcome, and examine whether experimental manipulations of memory support might play a role in subsequent treatment outcome. In such ways, the development of the MSRS provides a critical first step in harnessing a potentially powerful, generalizable, and readily disseminable mechanism of change – memory support.

Acknowledgments

This project was supported by National Institute of Mental Health Grant R34 MH094535. We are grateful to Dr. Rita Smith, Dr. Monique Thompson, Kerrie Hein, Niki Gumport, Dr. Steve Hollon, Dr. Kirsten Hamen, Dr. Joel Sherrill, Patrick Simpson, Anita Satish, Hanh Dang, Lynda Moy, Chris Fussell, and Dr. Joseph Williams for helpful input on the topic of this paper.

References

Airaksinen E, Larsson M, Forsell Y. Neuropsychological functions in anxiety disorders in population-based samples: Evidence of episodic memory dysfunction. Journal of Psychiatric Research. 2005; 39:207–214. DOI: 10.1016/j.jpsychires.2004.06.001 [PubMed: 15589570]

Almkvist O, Fratiglioni L, Agüero-Torres H, Viitanen M, Bäckman L. Cognitive support at episodic encoding and retrieval: Similar patterns of utilization in community-based samples of Alzheimer's disease and vascular dementia patients. Journal of Clinical and Experimental Neuropsychology. 2010; 21:816–830. DOI: 10.1076/jcen.21.6.816.862 [PubMed: 10649536]

Altman, DG. Comparing groups—categorical data. In: Altman, DG., editor. Practical statistics for medical research. 1. London, UK: Chapman and Hall; 1991. p. 261-265.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, D.C: American Psychiatric Association; 2000. Text revision

Bäckman L, Forsell Y. Episodic memory functioning in a community-based sample of old adults with major depression: utilization of cognitive support. Journal of Abnormal Psychology. 1994; 103:361–370. DOI: 10.1037/0021-843X.103.2.361 [PubMed: 8040505]

Bartlett MS. Tests of significance in factor analysis. British Journal of Psychology. 1950; 3:77–85. DOI: 10.1111/j.2044-8317.1950.tb00285x

Behnken A, Schöning S, Gerss J, Konrad C, de Jong-Meyer R, Arolt V. Persistent non-verbal memory impairment in remitted major depression - caused by encoding deficits? Journal of Affective Disorders. 2010; 122:144–148. DOI: 10.1016/j.jad.2009.07.010 [PubMed: 19692126]

Benson J. Developing a strong program of construct validation: A test anxiety example. Educational Measurement: Issues and Practice. 1998; 17:10–17. DOI: 10.1111/j.1745-3992.1998.tb00616.x

Bober SL, Hoke LA, Duda RB, Tung NM. Recommendation recall and satisfaction after attending breast/ovarian cancer risk counseling. Journal of Genetic Counseling. 2007; 16:755–762. DOI: 10.1007/s10897-007-9109-0 [PubMed: 17674165]

Bornstein RF. Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. Psychological Assessment. 2011; 23:532–544. DOI: 10.1037/a0022402 [PubMed: 21668126]

Bunce D. Cognitive support at encoding attenuates age differences in recollective experience among adults of lower frontal lobe function. Neuropsychology. 2003; 17:353–361. DOI: 10.1037/0894-4105.17.3.353 [PubMed: 12959501]

Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin. 1959; 56:81–105. DOI: 10.1037/h0046016 [PubMed: 13634291]

Carney RN, Levin JR. Pictorial illustrations still improve students' learning from text. Educational Psychology Review. 2002; 14:5–26. DOI: 10.1023/A:1013176309260

Cattell RB. The scree test for the number of factors. Multivariate Behavioral Research. 1966; 1:245–276. DOI: 10.1207/s15327906mbr0102_10 [PubMed: 26828106]

Chambers MJ. Patient recall of recommendations in the behavioural treatment of insomnia. Sleep Research. 1991; 20:222.

Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment. 1994; 6:284–290. DOI: 10.1037//1040-3590.6.4.284

Clark JM, Paivio A. Dual coding theory and education. Educational Psychology Review. 1991; 3:149–210. DOI: 10.1007/bf01320076

Coan, JA.; Gottman, JM. The specific affect coding system (SPAFF). In: Coan, JA.; Gottman, JM., editors. Handbook of emotion elicitation and assessment. New York, NY: Oxford University Press; 2007. p. 267-285.

Cohen, J. Statistical power analysis for the behavioral sciences. 2. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.

Craik FI, Govoni R, Naveh-Benjamin M, Anderson ND. The effects of divided attention on encoding and retrieval processes in human memory. Journal of Experimental Psychology: General. 1996; 125:159–180. [PubMed: 8683192]

Craik FI, Lockhart RS. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior. 1972; 11:671–684. DOI: 10.1016/s0022-5371(72)80001-x

Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychological Bulletin. 1955; 52:281–302. DOI: 10.1037/h0040957 [PubMed: 13245896]

DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, Gallop R. Cognitive therapy vs medications in the treatment of moderate to severe depression. Archives of General Psychiatry. 2005; 62:409–416. DOI: 10.1001/archpsyc.62.4.409 [PubMed: 15809408]

DeVellis, RF. Scale development: Theory and applications. 3. Thousand Oaks, CA: Sage Publications; 2012.

Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, Jacobson NS. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. Journal of Consulting and Clinical Psychology. 2006; 74:658–670. DOI: 10.1037/0022-006x.74.4.658 [PubMed: 16881773]

First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW., editors. Structured clinical interview for DSM-IV axis I disorders, clinician version (SCID-CV). Washington, DC: American Psychiatric Association; 2007.

Gielen S, Peeters E, Dochy F, Onghena P, Struyven K. Improving the effectiveness of peer feedback for learning. Learning and Instruction. 2010; 20:304–315. DOI: 10.1016/j.learninstruc.2009.08.007

Glorfeld LW. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. Educational and Psychological Measurement. 1995; 55:377–393. DOI: 10.1177/0013164495055003002

Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement: A meta-analysis of effects and processes. Advances in Experimental Social Psychology. 2006; 38:69–119. DOI: 10.1016/s0065-2601(06)38002-1

Gotlib IH, Joormann J. Cognition and depression: current status and future directions. Annual Review of Clinical Psychology. 2010; 6:285–312. DOI: 10.1146/annurev.clinpsy.121208.131305

Hart TJ. Memory and the memory-monitoring process. Journal of Verbal Learning and Verbal Behavior. 1967; 6:685–691. DOI: 10.1016/s0022-5371(67)80072-0

Harvey AG, Lee JY, Williams J, Hollon S, Walker MP, Thompson M, Smith R. Improving outcome of psychosocial treatments by enhancing memory and learning. Perspectives in Psychological Science. 2014; 9:161–179. DOI: 10.1177/1745691614521781

Henson RK, Roberts JK. Use of exploratory factor analysis in published research common errors and some comment on improved practice. Educational and Psychological Measurement. 2006; 66:393–416. DOI: 10.1177/0013164405282485

Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika. 1965; 30:179–185. DOI: 10.1007/bf02289447 [PubMed: 14306381]

Isaac CL, Cushway D, Jones GV. Is posttraumatic stress disorder associated with specific deficits in episodic memory? Journal of Psychiatric Research. 2006; 40:47–58. DOI: 10.1016/j.cpr.2005.12.004 [PubMed: 16199055]

Jansen J, Butow PN, Van Weert JCM, Van Dulmen S, Devine RJ, Heeren TJ, Tattersall MHN. Does age really matter? Recall of information presented to newly referred patients with cancer. Journal of Clinical Oncology. 2008; 26:5450–5457. DOI: 10.1200/jco.2007.15.2322 [PubMed: 18936478]

Kaiser HF. The application of electronic computers to factor analysis. Educational and Psychological Measurement. 1960; 20:141–151. DOI: 10.1177/001316446002000116

Kaiser HF. An index of factorial simplicity. Psychometrika. 1974; 39:31–36. DOI: 10.1007/BF02291575

Kapur N, Glisky EL, Wilson BA. Technological memory aids for people with memory deficits. Neuropsychological Rehabilitation. 2004; 14:41–60. DOI: 10.1080/09602010343000138

Karpicke JD, Roediger HL. The critical importance of retrieval for learning. Science. 2008; 319:966–968. DOI: 10.1126/science.1152408 [PubMed: 18276894]

Kolodner JL. Educational implications of analogy: A view from case-based reasoning. American Psychologist. 1997; 52:57–66. DOI: 10.1037//0003-066x.52.1.57 [PubMed: 9017932]

Kornell N, Castel AD, Eich TS, Bjork RA. Spacing as the friend of both memory and induction in young and older adults. Psychology and Aging. 2010; 25:498–503. DOI: 10.1037/a0017807 [PubMed: 20545435]

Ledesma RD, Valero-Mora P. Determining the number of factors to retain in EFA: an easy to use computer program for carrying out parallel analysis. Practical Assessment, Research and Evaluation. 2007; 12:1–11. DOI: 10.3758/bf03200806

Lee JY, Harvey AG. Memory for therapy in bipolar disorder. Journal of Consulting and Clinical Psychology. 2015; 83:93–102. DOI: 10.1037/a0037911

Lewkovich GN, Haneline MT. Patient recall of the mechanics of cervical spine manipulation. Journal of Manipulative and Physiological Therapeutics. 2005; 28:708–712. DOI: 10.1016/j.jmpt.2005.09.014 [PubMed: 16326241]

Ley P, Jain YK, Skilbeck CE. A method for decreasing patients' medication errors. Psychological Medicine. 1976; 6:599–601. DOI: 10.1017/s0033291700018237 [PubMed: 1005577]

Lobbestael J, Leurgans M, Arntz A. Interrater reliability of the structured clinical interview for DSM-IV axis I disorders (SCID I) and axis II disorders (SCID II). Clinical Psychology & Psychotherapy. 2011; 18:75–79. DOI: 10.1002/cpp.693 [PubMed: 20309842]

Martino DJ, Igoa A, Marengo E, Scarpola M, Strejilevich SA. Neurocognitive impairments and their relationship with psychosocial functioning in euthymic bipolar II disorder. The Journal of Nervous and Mental Disease. 2011; 199:459–464. DOI: 10.1097/nmd.0b013e3182214190 [PubMed: 21716059]

Mestre, JP., editor. Transfer of learning from a modern multidisciplinary perspective. Information Age Publishing; 2005.

Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. The Lancet. 2007; 370:851–858. DOI: 10.1016/s0140-6736(07)61415-9

Nunnally, JC. Psychometric theory. 2. New York, NY: McGraw-Hill; 1978.

O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instruments, and Computers. 2000; 32:396–402. DOI: 10.3758/bf03200807

Pickney CS, Arnason JA. Correlation between patient recall of bone densitometry results and subsequent treatment adherence. Osteoporosis International. 2005; 16:1156–1160. DOI: 10.1007/s00198-004-1818-8 [PubMed: 15744452]

Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): Psychometric properties. Psychological Medicine. 1996; 26:477–486. DOI: 10.1017/s0033291700035558 [PubMed: 8733206]

Rush AJ, Trivedi M, Wisniewski S, Nierenberg A, Stewart J, Warden D, Fava M. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR* D report. American Journal of Psychiatry. 2006; 163:1905–1917. DOI: 10.1176/ajp.2006.163.11.1905 [PubMed: 17074942]

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979; 86:420–3428. DOI: 10.1037//0033-2909.86.2.420 [PubMed: 18839484]

Siegel LL, Kahana MJ. A retrieved context account of spacing and repetition effects in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2014; 40:755–764. DOI: 10.1037/a0035585

Simpson HB, Maher MJ, Wang Y, Bao Y, Foa EB, Franklin M. Patient adherence predicts outcome from cognitive behavioral therapy in obsessive-compulsive disorder. Journal of Consulting and Clinical Psychology. 2011; 79:247–252. DOI: 10.1037/a0022659 [PubMed: 21355639]

Snyder HR. Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. Psychological Bulletin. 2013; 139:81–132. DOI: 10.1037/a0028727 [PubMed: 22642228]

Tabachnick, BG.; Fidell, LS. Using multivariate statistics. Needham Heights, MA: Allyn and Bacon; 2001.

Taconnat L, Baudouin A, Fay S, Raz N, Bouazzaoui B, El-Hage W, Ergis AM. Episodic memory and organizational strategy in free recall in unipolar depression: The role of cognitive support and executive functions. Journal of Clinical and Experimental Neuropsychology. 2010; 32:719–727. DOI: 10.1080/13803390903512645 [PubMed: 20155557]

Tosteson AN, Grove MR, Hammond CS, Moncur MM, Ray GT, Hebert GM, Ettinger B. Early discontinuation of treatment for osteoporosis. American Journal of Medicine. 2003; 115:209–216. DOI: 10.1016/s0002-9343(03)00362-0 [PubMed: 12947959]

Tulving E, Pearlstone Z. Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior. 1966; 5:381–391.

Vallis TM, Shaw BF, Dobson KS. The Cognitive Therapy Scale: psychometric properties. Journal of Consulting and Clinical Psychology. 1986; 54:381–385. DOI: 10.1037//0022-006x.54.3.381 [PubMed: 3722567]

Varga M, Magnusson A, Flekkoy K, David AS, Opjordsmoen S. Clinical and neuropsychological correlates of insight in schizophrenia and bipolar I disorder: does diagnosis matter? Comprehensive Psychiatry. 2007; 48:583–591. DOI: 10.1016/j.comppsych.2007.06.003 [PubMed: 17954145]

Videbech P, Ravnkilde B. Hippocampal volume and depression: a meta-analysis of MRI studies. American Journal of Psychiatry. 2004; 161:1957–1966. DOI: 10.1176/appi.ajp.161.11.1957 [PubMed: 15514393]

Widaman KF. Common factor analysis versus principal component analysis: Differential bias in representing model parameters? Multivariate Behavioral Research. 1993; 28:263–311. DOI: 10.1207/s15327906mbr2803_1 [PubMed: 26776890]

Williams JJ, Lombrozo T. The role of explanation in discovery and generalization: evidence from category learning. Cognitive Science. 2010; 34:776–806. DOI: 10.1111/j.1551-6709.2010.01113.x [PubMed: 21564236]

Williams JJ, Lombrozo T, Rehder B. The hazards of explanation: Overgeneralization in the face of exceptions. Journal of Experimental Psychology: General. 2013; 142:1006–1014. DOI: 10.1037/a0030996 [PubMed: 23294346]

Williams RM, Moorey S, Cobb J. Training in cognitive-behaviour therapy: Pilot evaluation of a training course using the Cognitive Therapy Scale. Behavioural Psychotherapy. 1991; 19:373–376. DOI: 10.1017/s0141347300014075

Yoo SS, Hu PT, Gujar N, Jolesz FA, Walker MP. A deficit in the ability to form new human memories without sleep. Nature Neuroscience. 2007; 10:385–392. DOI: 10.1038/nn1851 [PubMed: 17293859]

Young, J.; Beck, AT. Unpublished manuscript. University of Pennsylvania; Philadelphia, PA: 1980. Cognitive therapy scale: Rating manual.

Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. Psychological Bulletin. 1986; 99:432–442. DOI: 10.1037/0033-2909.99.3.432

Appendix

Appendix A

The Memory Support Rating Scale (MSRS)

Memory Support Strategies	Tally
Attention Recruitment. Involves the treatment provider 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 ten years time, it is this skill” or “this is a key point to remember”), or multimedia/diverse presentation modes (e.g., handouts, poems, songs, note taking, role-playing, imagery, using a white board) as a means to recruit the patient’s attention.	
Categorization. Involves explicit effort by the treatment provider 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. Involves the treatment provider 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 all day when you are feeling depressed?”).	
Application. Involves the treatment provider 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?”).	
Repetition. Involves the treatment provider restating, rephrasing, or revisiting information discussed in treatment (e.g., “in other words,” “as we talked about earlier,” or “in sum”).	
Practice Remembering. Involves the treatment provider 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?).	

Memory Support Strategies	Tally
Cue-Based Reminder. Involves the treatment provider helping the patient develop new or existing cues (e.g., colored wrist bands, reminder text messages/phone calls/e-mails, smart phone apps, acronyms, rhymes, and other mnemonics) to facilitate memory for treatment points.	
Praise Recall. Involves the treatment provider 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.").	

MSRS Scores:

Memory Support Instances (sum of tallies): _________

Memory Support Items Used (out of 8): _________

Memory Support Instances per Minute: _________

Table 1

Participant Characteristics

	CT+MS (*n* = 22)	CT-as-usual (*n* = 20)	Test Statistic (df)	*p*
Gender, *n* (% Female)	9 (40.9)	14 (70.0)	$\chi^2(1) = 3.58$	.059
Race				
White, *n* (%)	18 (81.8)	13 (65.0)		
Black, *n* (%)	0	1 (5.0)		
Asian, *n* (%)	1 (4.5)	3 (15.0)		
Native American, *n* (%)	1 (4.5)	0		
Multi-Racial, *n* (%)	0	1 (5.0)		
Not Specified, *n* (%)	2 (9.1)	2 (10.0)		
			$\chi^2(5) = 4.72$	.451
Ethnicity				
Non-Hispanic, *n* (%)	18 (81.8)	14 (70.0)		
Hispanic, *n* (%)	4 (18.2)	3 (15.0)		
Not Specified, *n* (%)	0	3 (15.0)		
			$\chi^2(2) = 3.56$	.169
Age (years)	42.18 (8.10)	45.25 (12.02)	$t(32.85) = -0.98$	.334
Education Level (years)	15.59 (1.59)	16.53 (1.98)	$t(39) = -1.68$	.102
IDS-SR (Pre)	38.82 (8.24)	42.55 (9.55)	$t(40) = -1.36$	.182

Note. Mean (SD) presented unless otherwise noted. MSRS = Memory Support Rating Scale; CT+MS = Cognitive Therapy + Memory Support; CT-as-usual = Cognitive Therapy as Usual; IDS-SR = Inventory of Depressive Symptomatology − Self-Report.

Table 2

Inter-Item Correlations of the Memory Support Rating Scale

MSRS Items	1	2	3	4	5	6	7	8
1. Attention Recruitment	--							
2. Application	.40*	--						
3. Evaluation	.39*	.36*	--					
4. Categorization	.30*	.03	.33*	--				
5. Repetition	.64*	.45*	.46*	.31*	--			
6. Practice Remembering	.49*	.50*	.33*	.19	.64*	--		
7. Cue-Based Reminder	.23	.28*	.12	.17	.26	.25	--	
8. Praise Recall	.33*	.43*	.32*	.05	.29*	.52*	.25	--

Note. MSRS = Memory Support Rating Scale. Inter-item correlations are presented for $n = 171$ tapes coded.

* $p < .001$.

Table 3

Means of Validity Measure Scores by Treatment Condition

Construct Validity Measure	CT+MS (n = 22)	CT-as-usual (n = 20)	t	df	p	d
Group Differentiation Ability						
MS Instances	18.57 (8.95)	8.39 (3.90)	4.70	29.26	< .001	1.45
MS Items Used	4.93 (1.11)	3.40 (0.70)	5.27	40.00	< .001	1.63
MS Instances per Minute	0.30 (0.14)	0.14 (0.06)	4.83	30.25	< .001	1.46
Predictive Validity						
Patient Recall Task	9.86 (6.13)	7.94 (3.37)	1.26	33.70	.218	0.38
Discriminant Validity						
CTRS	47.52 (4.87)	47.29 (4.08)	0.16	40.00	.871	0.05

Note. Mean (SD) presented. MSRS = Memory Support Rating Scale; CT+MS = Cognitive Therapy + Memory Support; CT-as-usual = Cognitive Therapy as Usual; MS = Memory Support; CTRS = Cognitive Therapy Rating Scale.

Table 4

Correlations between MSRS Scores and Validity Measure Scores

Validity Measure	MS Instances	MS Items Used	MS Instances per Minute
Convergent Validity			
Patient Recall Task	$r = .29$	$r = .36$	$r = .32$
($n = 40$)	$p = .073$	$p = .022$	$p = .044$
Discriminant Validity			
CTRS	$r = .07$	$r = .13$	$r = .08$
($n = 42$)	$p = .674$	$p = .424$	$p = .615$

Note. Mean MSRS Scores across randomly selected sessions per participant are presented. MSRS = Memory Support Rating Scale; MS = Memory Support; CTRS = Cognitive Therapy Rating Scale.