Internet Interventions 18 (2019) 100259

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent

Efficacy of online Memory Specificity Training in adults with a history of depression, using a multiple baseline across participants design

Kris Martens[a], Tom J. Barry[b,c,*], Keisuke Takano[d], Patrick Onghena[a], Filip Raes[a]

[a] Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium
[b] Department of Psychology, The University of Hong Kong, Hong Kong, Hong Kong
[c] Department of Psychology, The Institute of Psychiatry, King's College London, London, United Kingdom
[d] Ludwig-Maximilians-University of Munich, Munich, Germany

ARTICLE INFO

Keywords:
Memory Specificity Training
Autobiographical memory
Multiple baseline design
Depression

ABSTRACT

Background and objectives: Memory Specificity Training (MeST), a group training protocol, is effective in improving autobiographical memory specificity (AMS), and in so doing, reducing emotional disorder symptoms amongst clinical groups. We examined MeST's effectiveness when the core component (memory specificity trials) is offered online and individually (c-MeST).

Methods: A multiple-baseline across-participants design with a randomization-to-baseline length (14 to 33 days) was used. Participants were twenty adults (16 female; M_{age} = 50, SD = 12) experiencing reduced AMS, at least one lifetime depressive episode and who currently reported at least minimal depressive symptoms. During baseline, the training phase (nine sessions across 17 days) and a three-month follow-up assessment, AMS, depressive symptoms and related processes were measured.

Results: AMS improved significantly by three months follow-up. Session-to-session scores indicated that AMS improved most from baseline to the first online session, with no further improvement thereafter. In contrast to studies with clinical participants, no significant change in symptoms or secondary processes such as rumination was found.

Conclusions: Translating MeST into an online, individual version is a feasible, low-cost intervention for reduced AMS. Future research should examine c-MeST's potential for preventing increases in symptoms in at-risk samples with longer follow-ups as well as its potential for reducing symptoms in clinical groups.

1. Introduction

As depression continues to be the leading cause of disability worldwide (World Health Organization, 2017), there is a critical need for accessible, low-cost interventions. One possible route to meet this need is to link risk factors with specific therapeutic interventions (Craske, 2018). The current study translates one such intervention, Memory Specificity Training (MeST) to an online scalable application.

There is substantial variability between people in their ability to retrieve specific, personal memories of events lasting less than a day. Such difficulty is referred to as reduced Autobiographical Memory Specificity (rAMS) or Overgeneral Autobiographical Memory (Williams et al., 2007). For example, given the cue word "Animal" the response "*I remember the day we got our first cat*" refers to a specific autobiographical memory, whereas someone with rAMS may answer such a question with a categorical event that took place repeatedly (e.g. "*I always cuddled with our cat*") or an extended memory of an event lasting longer than one day (e.g. "*That time we were travelling in Turkey and saw a lot of cats*"). It is assumed that autobiographical memories are retrieved by searching autobiographical knowledge which is hierarchically ordered from general, summarized information to event-specific details (Conway and Pleydell-Pearce, 2000). The top-down search through this hierarchy can be disrupted by several mechanisms, described within the CaR-FA-X model (Williams, 2007): Capture and Rumination, Functional Avoidance, and impaired eXecutive Control.

rAMS can be regarded as a clinically relevant phenomenon due to its association with a range of psychological processes and outcomes. For instance, it is associated with an increase in rumination (Nolen-Hoeksema, 2000; Starr and Davila, 2012), an impairment in interpersonal (Raes et al., 2005; Sutherland and Bryant, 2008), and non-interpersonal problem solving (Beaman et al., 2007; Goddard et al., 1996), a generalized and negative view of one's self and the world

* Corresponding author at: Department of Psychology, The University of Hong Kong, Jockey Club Tower, Pokfulam Road, Hong Kong.
E-mail address: tom.j.barry@icloud.com (T.J. Barry).

https://doi.org/10.1016/j.invent.2019.100259
Received 1 February 2019; Received in revised form 26 June 2019; Accepted 29 June 2019
Available online 15 July 2019
2214-7829/ © 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

(Watkins et al., 2009), an impairment in imagining future events (Jing et al., 2016) and hopelessness about one's future (Raes et al., 2009b), and impairments in coping with distressing memories and feelings (Debeer et al., 2012).

Given these associations, it is no wonder that rAMS is considered to be an enduring trait of depression (Williams et al., 2007) and a predictor of depressive symptoms at follow-up over and above initial symptom levels (Sumner et al., 2010). Even when depressed patients are recovered or in remission, specificity does not improve (Mackinger et al., 2000; Park et al., 2002). In addition, rAMS can be regarded as a transdiagnostic factor, as it is also found in Anorexia Nervosa (i.e. Bomba et al., 2014; Huber et al., 2015), Bipolar Disorder (i.e. Tzemou and Birchwood, 2007), Post-Traumatic Stress Disorder (i.e. McNally et al., 1995) and is also associated with aging regardless of the presence of emotional disorders (i.e. Wilson and Gregory, 2017).

In the past decade, interest has increased in the potential of autobiographical memory training in the treatment of emotional disorders and particularly depression (Hitchcock et al., 2017). There is support for the hypothesis that rAMS is modifiable when targeted through intervention, and that this modification has a concomitant effect on the symptoms of emotional disorder and other secondary processes that mediate the association between rAMS and disorder (Barry et al., 2019). A first uncontrolled clinical trial examined the effect of a four-session group training called Memory Specificity Training (MeST); it reduced rAMS and ameliorated a number of associated cognitive processes (improvements in problem solving, rumination and hopelessness) in 10 depressed female inpatients (Raes et al., 2009b). More recently, in another uncontrolled clinical trial, 26 depressed outpatients showed an increase in memory specificity and a decrease in depressive symptoms (Eigenhuis et al., 2017). A randomized controlled pilot trial of MeST versus a no-treatment control of 23 bereaved, depressed Afghani teenagers in Iran (Neshat-Doost et al., 2013) also showed an increase in memory specificity compared with wait-list controls. The effects of MeST on depressive symptoms were delayed in this study such that there was no group difference by post-intervention assessment but for MeST participants' symptoms continued to improve after treatment and were significantly lower than controls by two-month follow-up assessment. A cluster-randomized controlled platform pilot trial investigating MeST relative to Psychoeducation and Supportive Counselling (PSC) for Major Depressive Disorder ($N = 62$) also showed improvement in memory specificity at post-intervention and follow-up for the MeST-group relative to the PSC (Werner-Seidler et al., 2018). A reduction in both groups in depressive symptoms was found at a 3-month follow-up. A recent meta-analysis and systematic review (Barry et al., 2019) concluded that MeST is associated with improvement in memory specificity and depressive symptoms, but the benefits did not last until follow-up assessment.

The core component of MeST resembles the Autobiographical Memory Test (AMT, Williams and Broadbent, 1986), which assesses rAMS by presenting participants with cue words and instructing them to retrieve specific memories which these cue words remind them of. In MeST, participants receive the same instructions with positive, negative and neutral cue words. In addition, in a second kind of specificity exercises they are instructed to retrieve specific memories of the past day (with no cue words given). After retrieving a specific memory, participants are encouraged to retrieve details of this specific moment. In the first study examining MeST, participants were offered a total of 104 such retrieval trials (Raes et al., 2009b). In addition to these specificity trials, MeST protocols consist of (a) psycho-education on memory problems in depression; (b) psycho-education and exercises on how to notice when one is thinking at an overgeneral level and how to counter this maladaptive thinking by switching to a more specific one; and, (c) therapist and group interactions. The degree of difficulty of specificity trials increases stepwise in MeST by (a) increasing the amount of specificity trials per session; (b) by changing the valence of the cue words (negative cue words are only introduced in session three); and, (c) by adding more complex trials later in the training (e.g. finding two memories for two opposing cue words like skillful and clumsy).

However, several questions remain regarding the mechanism by which MeST exerts its effects on memory specificity and the symptoms of pathology. First, it remains unclear if the specificity trials in MeST are sufficient to produce change in memory specificity, even in the absence of the psychoeducational aspects of MeST. Second, it is also unclear whether this stepwise approach of increasing difficulty is necessary to train autobiographical memory specificity or whether this contributes towards MeST's feasibility or acceptability with participants. Finally, it is also unclear *how much* training is required in order to modify memory specificity within MeST. Throughout the different studies, MeST-protocols extended from four (Raes et al., 2009b) to five sessions (Eigenhuis et al., 2017), however, it is unclear how many sessions or trials are necessary.

Given that MeST is based on similar specificity trials as the AMT, a recently designed computerized scoring algorithm for scoring AMT responses (Takano et al., 2017c) offers new possibilities given that MeST might now be delivered in the absence of a therapist. Takano et al. (2017b) designed and examined an online version of MeST (named c-MeST) using the computerized scoring algorithm (Takano et al., 2017c; Takano et al., 2018). In a proof of concept study involving participants with rAMS (AMT score $<$ 50%), c-MeST effectively reduced rAMS after a two-week training (consisting of seven sessions of each 5 to 8 trials) and at a two-week follow up measurement, compared to a no-training control group. However, no effect on depressive symptoms was found post-training and follow-up. The authors anticipated that they not find a direct effect on secondary outcomes based on previous finding (Neshat-Doost et al., 2013) in which the impact of MeST on depressive symptoms was only visible at a two-month follow-up. For this reason, the current study will include a three-months follow-up using repeated measures.

These results imply that rAMS can be modified by only using specificity trials, but due to the design of the training do not answer questions regarding how much training is required and if a stepwise approach of increasing difficulty is necessary to train MeST. In addition, it is unclear how participants experience the online training and computerized feedback.

This study examines an online version of MeST that exclusively uses specificity trials without a gradual increase in difficulty using the computerized scoring algorithm. Moreover, in this version of online MeST (c-MeST) sessions are standardized, resulting in session-by-session specificity scores.

In addition to pre-intervention and follow-up measurement of memory specificity to assess whether c-MeST increases memory specificity and session-by-session scores to gain insight in how much training is needed, a multiple baseline single-case experimental design (SCED) is used to examine the impact of c-MeST on repeatedly measured symptoms and related processes. This design has several advantages; due to the frequency of the measurements one can observe if and when change happens, and to what degree (Hayes, 1981). The randomized multiple baseline component also allows control over known and unknown time-related confounding variables (Heyvaert and Onghena, 2014).

The principle aim of this study was to investigate whether an online, individually-delivered version of Memory Specificity Training (referred to as c-MeST) that exclusively consists of specificity trials leads to an increase in memory specificity in people who showed rAMS and experienced at least one depressive episode and reported at least mild depressive symptoms. For the current study, an inclusion criterion of 70% on the AMT is used as we agree with Werner-Seidler et al. (2018) that a deficit in memory specificity was a necessary criterion to be sure that there was opportunity to benefit from c-MeST. The secondary inclusion criterion of reporting at least one depressive episode in the past (assessed with the Major Depressive Questionnaire) and at least reporting minimal depressive symptoms (assessed with the Patient Health Questionnaire) was used because this combination increased the chance

that the current study examined the potential impact of c-MeST in a sample at risk for depression, as depression is a highly recurrent disorder (Burcusa and Iacono, 2007).

We expected an increase of memory specificity between pre-intervention and a three months follow-up measurement. A second research question concerned whether c-MeST impacted depressive symptoms, plus related processes and symptoms: (1) rumination, (2) worry, (3) unwanted thoughts or images, (4) being tense when a painful memory arises, and (5) thought suppression. In addition, we assessed happiness and sadness.

We expected a significant decrease between baseline (phase A) and training and follow up phase (phase B) for depressive symptoms and related symptoms, states (sadness, being tense when a painful memory arose) and processes (rumination, worry, thought suppression), and an increase for happiness. A third research question was if c-MeST is a feasible option to remediate rAMS; did participants complete c-MeST and how did they experience the cue words and provided feedback? An additional fourth research question examined how much training was necessary to observe a change in memory specificity.

2. Method

2.1. Participants

Participants were 20 individuals (16 female) who met inclusion criteria: (a) experiencing rAMS, operationalized as a score of < 70% on the Autobiographical Memory Test (Williams and Broadbent, 1986), (b) having experienced a Major Depressive episode, as measured with the Major Depressive Questionnaire (MDQ), (c) suffering from at least mild depressive symptoms operationalized as a score of at least 5 out of the maximum 27 on the Patient Health Questionnaire ($M = 12.4$, $SD = 6.01$) and (d) being in contact with a clinician in case their psychological condition would worsen; being in follow-up or in treatment with a clinical psychologist ($n = 14$; 70%) or having informed their general practitioner about their psychological condition ($n = 6$; 30%). None of the participants were inpatients. A flow diagram of the selection and inclusion process is illustrated in Fig. 1.

Participants' ages ranged between 27 and 74 ($M_{age} = 50$, $SD = 12$) years. Half of the sample had a full-time paid job, the other were unemployed (retired, sick-leave, never worked, volunteering). For the marital status, 75% of participants had a partner (married, living together), 25% were single. Nine (45%) participants reported a current Major Depressive episode. Sixteen participants (80%) suffered from recurrent depressive episodes. The average age of onset of participants' first depressive episodes varied between 10 and 70 years ($M = 27.37$, $SD = 15.57$).

2.2. Measures

2.2.1. Autobiographical Memory Specificity

Autobiographical Memory Specificity was measured using a verbal version of the Autobiographical Memory Test with two sets of cues for pre-training and follow-up assessments (Dalgleish et al., 2007; Williams and Broadbent, 1986). Ten cue words (five positive, five negative, presented in alternating order, presented in Appendix, Table A.1) were verbally presented. In response to each cue word, participants were instructed to retrieve a specific memory. It was explained that the memory does not have to be an important event, but needs to be specific, happened once and lasted shorter than a day. Within the instructions, examples of specific and non-specific responses were given and a practice trial with three cue words with feedback took place. The AMT was scored during the verbal interview by the first author, unspecific and unclear answers were followed with prompts until participants succeeded in retrieving a specific answer or until one minute passed. The AMT was administered at pre-intervention and follow-up assessments for which the two cue sets (AMT A and AMT B) were counterbalanced; half of the participants were offered the A version at pre measurement and the B version at follow-up measurement and half of the participants were offered them the other way round. AMT is scored as the number of cue words for which participants' first answer is classified as a specific autobiographical memory. Previous studies varied in the AMT scores that thy used as inclusion criteria, from no inclusion criterion (Eigenhuis et al., 2017; Neshat-Doost et al., 2013; Raes et al., 2009b) to scoring lower than 70% (Werner-Seidler et al., 2018) to scoring lower than 50% (Takano et al., 2017b). For the current study, an inclusion criterion of 70% on the AMT is used. As a point of reference, pre-intervention specificity scores in studies examining the impact of MeST using no inclusion criterion are 44% (Raes et al., 2009b), 63% (Eigenhuis et al., 2017), 63% and 61% (Neshat-Doost et al., 2013).

2.2.2. Depressive symptomatology, current and past major episodes

The Patient Health Questionnaire 9 (Kroenke et al., 2010) was used to measure depressive symptomatology. The PHQ-9 is a nine-item self-report measure of depressive symptoms which refers to DSM IV depression diagnostic criteria and other leading Major Depressive Disorder symptoms. The PHQ-9 was used as an inclusion criterion, for which participants needed to report minimal symptoms (> 5) for the previous two weeks, as this is regarded as minimal depressive symptoms (Kroenke et al., 2001). But PHQ-9 is also used as a repeated measurement during the study (baseline, intervention phase and follow up phase), to measure change within the individual, applied to the last day. Scores can vary from 0 to 27. PHQ-9 showed good internal consistency ($\alpha = 0.85$ for the first of the repeated measurements).

The Major Depression Questionnaire (MDQ; Van der Does et al., 2003) is a self-report measure for screening for current and past Major Depressive episodes asking participants about DSM IV criteria for current and past major depression, also including questions on functioning and exclusion criteria. MDQ was used as a pre measurement to assess past and current major depressive episodes, number of past depressive episodes and age of onset of a first depressive episode.

2.2.3. Rumination

The Ruminative Response Scale – Brooding subscale (RRS Brooding; Raes et al., 2009a; Treynor et al., 2003) was used to assess change in the brooding subtype of depression across the intervention. The RRS Brooding is a self-report questionnaire consisting of five items measuring brooding, part of the 22 items of the Ruminative Response Scale (RRS, Nolen-Hoeksema and Morrow, 1991) and shows a good internal consistency ($\alpha = 0.83$ for the first of the repeated measures). The items on the brooding factor are considered to measure the maladaptive coping of passively comparing one's situation with some unachieved standard. E.g., participants are asked to report how frequently they tend to think “Why do I always react this way?” or “Why do I have problems other people do not have?” on a 1 (*almost never*) to 4 (*always*) scale. Questions are applied to the last day (“since yesterday”). Scores vary from 5 to 20. The RRS Brooding was used as a part of the repeated measures.

2.2.4. Items for repeated measurements

A multiple baseline across participants design requires repeated measures. Depressive symptoms and rumination were measured repeatedly using the PHQ-9 and RRS Brooding. In addition, 8 single items were added that capture change in symptoms and related psychological processes: (1) rumination, (2) worry, (3) unwanted thoughts or images, (4) being tense when a painful memory arises, and (5) thought suppression. In addition, we assessed happiness and sadness. Participants scored to what extent they experienced, within the last day, each of the situations referred to within the eight statements mentioned on Likert scales from 1 (*not at all*) to 9 (*almost all of the time*): (a) “I worried about the future”, (b) “I worried about the past”, (c) “Unwanted images or thoughts that suddenly arose bothered me”, (d). “I tried to ban

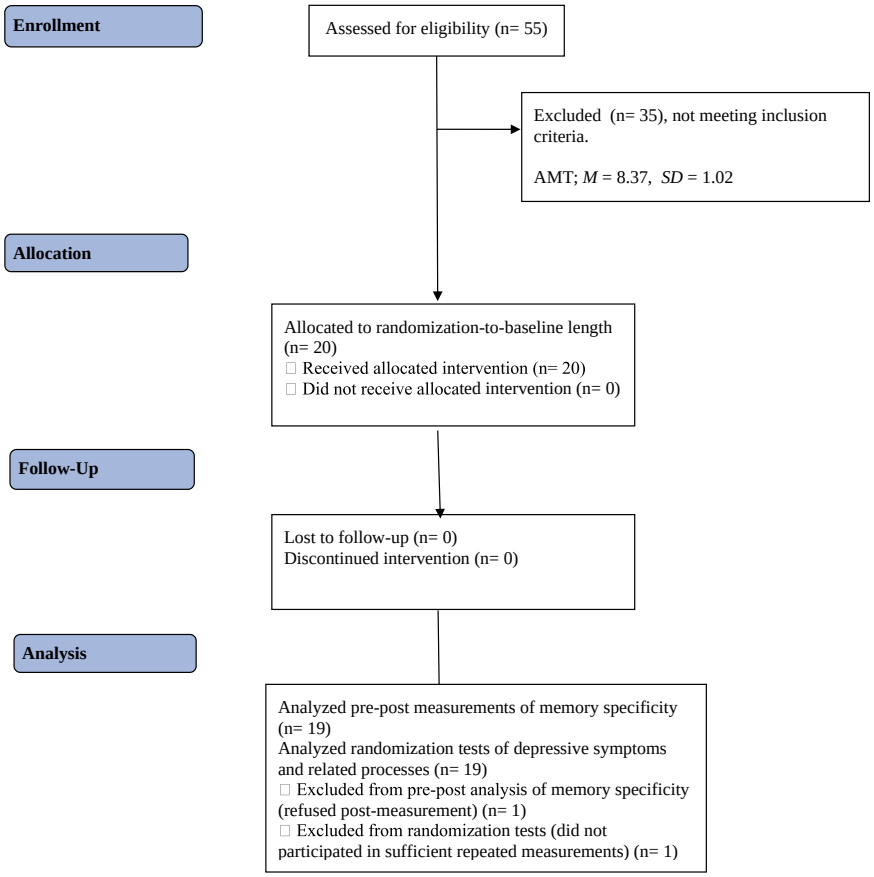


Fig. 1. A flow chart of the selection and inclusion process.

unwanted images or thoughts", (e) "If a painful memory arose, I got tense", (f) "I felt sad", (g) "I felt happy", (h) "I noticed the training had a positive impact on my functioning". Items (c), (d) and (e) are combined for analysis, as they are inspired by the three subscales of the revised Impact of Events scale (Weiss and Marmar, 1995).

2.2.5. Measures of training experiences

After each session participants are asked three closed and two open questions, asking participants (a) to what extent they found the offered words helpful/easy to help them retrieve a specific memory (0 = *not easy at all, words are very difficult to retrieve memories for*, 10 = *very easy, words are very easy to retrieve memories for*), (b) to what extent they experienced the feedback of the classifier as correct (0 = *not at all, a lot of mistakes*, 10 = *very correct, no mistakes*), (c) to what extent they experienced the session okay in length (1 = *way too short*, 2 = *a bit too short*, 3 = *just right*, 4 = *a bit too long*, 5 = *way too long*), (d) how they experienced the training and (e) if they had any other remarks.

2.3. The intervention - online Memory Specificity Training (c-MeST)

Online MeST consists of nine sessions of eleven specificity trials. In each specificity trial, participants are asked to retrieve a specific memory. Immediate feedback on their generated memory is provided. Online MeST is provided via the internet. Participants complete each session on an online platform which contained instruction and tips about autobiographical specificity, similar to the instructions of the AMT but more examples are provided.

Sets of cue words were created by first selecting words from a corpus of 4300 Dutch words (Moors et al., 2013) on the basis of the acquisition age and emotional valence. Words with an acquisition age lower than 6 were selected and the valence scores for each word varied from 1 (very negative) to 7 (very positive). Nine sets of words were created by dividing negative, positive and neutral (that is, those with a score of four in the aforementioned valence scale) words one by one into different sets, until each set contained three negative, three positive and three neutral cues. These 9 sets were checked on valence/pleasantness, activity/arousal, power/dominance (Moors et al., 2013) and concreteness (Brysbaert et al., 2014).

Similar to the structure of therapist-provided MeST, each session the nine trials with cue words are followed up by one trial about a memory of yesterday and one about today (without cue words). The website uses the computerized scoring algorithm for the Autobiographical Memory Test (Takano et al., 2017c) to score entries and to automatically give feedback if the entry is specific or not. The scoring algorithm showed good performances against expert-rated scores in discriminating specific versus non-specific memories (> 0.90 as Area under the Curve in Receiver Operating Characteristic analysis; Takano et al., 2018). If the entry is not specific enough, participants are encouraged to re-enter the memory (or another memory) with more episodic details. For each cue word, participants get three chances to enter a specific memory and if they cannot generate a specific memory within three attempts the next cue word is presented automatically. If participants succeed in providing a specific memory, they are then invited to provide more spatio-temporal and contextual details on the next page (i.e., "Where did it happen? When did it happen? How long did it take? Who else was there? What can u see, hear, smell or taste? What kind of day was it?"). Participants are instructed to only fill out these details if they did not already provide these details in their successful entry. Participants can skip a cue word if they wish to do so. There is no time limit per question. In sum, participants are offered nine sessions of eleven specificity trials in 17 days, which is a similar in dose as original MeST (99 specificity trials versus 104 specificity trials in 4 weekly sessions and as

homework assignments between sessions; Raes et al., 2009a, 2009b) but higher in dose as the previous examination of c-MeST which did not impacted secondary measures (Raes et al., 2009a, 2009b).

Entries were scored manually by the first author. c-MeST sessions are scored as the number of trials for which participants' first answer is classified as a specific autobiographical memory, in line with the AMT, resulting in a maximum of 11 points per session.

2.4. Procedure

To examine the efficacy of an online version of the Memory Specificity Training (c-MeST), a multiple baseline across participants design with a fixed number of participants ($n = 20$) was chosen. The randomized multiple baseline component allows control over known and unknown time-related confounding variables by varying the length of the baseline (intervention randomization). A minimum baseline of two weeks (14 days) was chosen to observe potential variation in participants before the intervention phase started. Using randomization tests 20 randomized baselines are necessary to create the option of a minimal p value of 0.05 per participant. Regarding statistical power, with a minimum number of measures of 14 per phase, a minimum of 62 measures in total for the participant with the shortest baseline, and 20 participants, previous simulation studies indicate that this study can reach sufficient statistical power (Heyvaert et al., 2017; Michiels et al., 2018). The 20 baselines varying from 14 to 33 days were randomized and then allocated to participants who met inclusion criteria. This process stopped after the preset sample size (20 participants) was met. The research design is illustrated in Fig. 2.

Participants were recruited between September 2016 and July 2017 via two local Flemish newspapers and a local television station, in which prospective participants were informed that the first results of MeST as an add-on intervention for depression are promising in improving specificity and depressive symptoms. Prospective participants were invited for a meeting in which they received additional study information after which they gave informed consent. Those who met inclusion criteria received a personalized file with a schedule of which day they had to train and/or fill out questionnaires, based on the allocated randomized baseline length (intervention randomization) and starting the day after the meeting (start-point randomization). The 20 possible baselines were randomized using random.org and then assigned to participants in the order of intake meetings. Each day when participants were asked to fill out a questionnaire or training, an invitation was sent by e-mail (automated with the software Boomerang) to remind them. On a day when participants needed to fill out both, participants received both e-mails at the same time. We did not instruct participants in which order the training and questionnaire should be completed. An online platform was created where participants could fill out the repeated measurements throughout the study: PHQ-9, RRS Brooding and single items.

Before the training phase started, participants received a phone call wherein the instructions about the training were repeated. After the training phase participants received an e-mail with instructions on how to train further in their own time with pen and paper trials, if they wished to. After the follow-up finished, participants were invited for a post-intervention assessment in person or by telephone. No blinding was used. The study received institutional ethical approval of the Social and Societal Ethics Committee of KU Leuven.

2.5. Analysis

Scores of the pre- and follow-up measurement of AMT were submitted to a t-test to confirm whether c-MeST leads to an increase in autobiographical memory specificity. Session-by-session scores, which are manually scored by one of the authors, are used to explore how much training is needed. To measure the impact of c-MeST on depressive symptoms and related states (sadness, happiness, being tense when a painful memory arose) and processes (rumination, worry, thought suppression), a multiple baseline across participants design is used and a visual and statistical analysis are applied. The baseline length (with repeated measurements every day), phase A, is randomized between 14 and 33 days. We did not have a clear hypothesis about a change point where a (clinically) significant effect on secondary measures could be observed; in some studies an effect was found at both post-intervention measurement and follow-up (Eigenhuis et al., 2017; Werner-Seidler et al., 2018), but Neshat-Doost et al. (2013) only found an effect on secondary measures at the follow-up measurement. For this reason, we statistically compared the randomized baselines with a combination of the training and follow-up phase to detect an impact. The training phase (17 days, with measurements each day) and follow-up phase (3 months, with measurements every three days) are combined as phase B. This results in a minimum of measurement times per phase of 14, and a total number of measurement times varying from 62 to 82. Although the training phase and follow-up phase are combined for statistical analysis, a post-hoc visual analysis (according to the criteria defined by Kratochwill et al., 2010) was applied to check whether an effect occurred during the treatment that would have disappeared during follow-up, or that a clear effect appeared only during follow-up. For visual and statistical analysis the software 'Single Case Data Analysis' (Bulté and Onghena, 2013) is used, which uses the R packages SCRT, SCVA and SCM. For visual analysis with SCVA, the level of each phase using mean scores and the trend of each phase using least squares regression are examined (Bulté and Onghena, 2011). For statistical analysis, randomization tests (RTs) are calculated. Randomization tests are statistical significance tests based on the random assignment of experimental units to treatments (Onghena and Edgington, 2005). First, a test statistic (TS) is calculated. For all dependent variables a decrease in scores is expected, and thus TS was defined as B – A; one exception is the item, happiness, for which an increase in scores is expected and thus TS was defined as A – B. Second, the TS for each possible assignment ($N = 20$) is calculated given the collected data of the participant. Third, we examined where the obtained TS falls within the distribution of all possible test statistic values. The p value of the RT is calculated as the proportion of possible TS values that is as extreme, or even more extreme, than the value of the TS based on the collected data. And thus, with 20 participants, the minimum possible p value is 0.05 per participant. These p values are then combined in a multiplicative way (Onghena and Edgington, 2005).

3. Results

3.1. Compliance and missing values

One participant declined the follow-up assessment of autobiographical memory specificity (AMT) and one participant did not report sufficient repeated measurements to be included for the

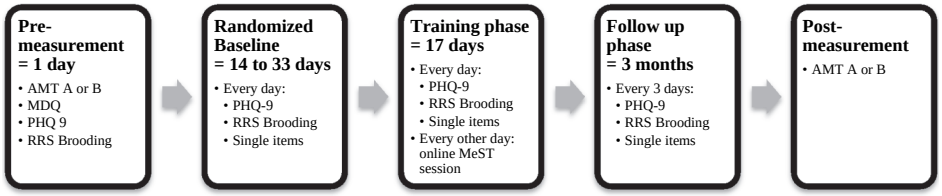


Fig. 2. Visualized graphic illustration of the research design per participant, AMT = Autobiographical Memory Test, MDQ = Major Depression Questionnaire, RRS Brooding = Ruminative Response Scale Brooding Subscale, MeST = Memory Specificity Training.

Table 1
Minima, maxima, means and standard deviations of amount of session and trials made, and scores for total amount of trials as well as for only negative, neutral as positive cues as memories of the day.

	Min	Max	M	SD
Number of sessions	7	9	8.65	0.57
Number of trials	72	99	93.7	8.09
Total score %	41.41	100.00	80.63	13.87
Number of trials with negative cues	19	27	25.65	2.24
Score (%) of trials with negative cues	25.93	100.00	73.38	19.12
Number of trials with neutral cues	19	27	25.45	2.40
Score (%) of trials with neutral cues	33.33	100.00	79.69	16.98
Number of trials with positive cues	20	27	25.6	2.13
Score (%) of trials with positive cues	33.33	100.00	76.74	18.33
Number of trials with memories of the day	14	18	17	1.45
Score (%) of trials with memories of the day	88.89	100.00	98.25	3.21

randomization tests.

Overall participants showed a good compliance, completing 94.65% of the trials (SD = 8.17) and an average score of 80.63% (SD = 13.87%) of the intervention trials. For offering twenty participants nine sessions each with eleven trials, the mean amount of sessions completed was 8.65 (SD = 0.57). Table 1 shows the descriptive statistics for participants' performance on each trial.

3.2. Effects of c-MeST on memory specificity during training and at three months follow up

Participants' scores on the AMT increased significantly, $t(18) = -6.94$, $p < .001$, from pre- (M = 4.85, SD = 1.66) to post-intervention assessment measurement after three months follow up (M = 7.79, SD = 1.99). Between the last session and the three months follow up participants were able to train with provided pen and paper trials; 14 people (70%) never did, 2 (10%) tried less than ten trials and 4 participants (20%) made between ten and fifty extra trials. Table A.2 in the Appendix shows for each participant specificity scores for pre-intervention and follow-up measurement, and for each session. Fig. 3 illustrates pre-intervention and follow-up measurement AMT mean scores and mean scores of the group for each session which varies from 75.86% to 86.59% with an overall mean score of 80.99%. Visual analysis of session-to-session scores shows that there was a significant improvement in the proportion of specific answers given by participants from the pre-intervention assessment to the end of the first session, after which no further enhancement of specificity was observed.

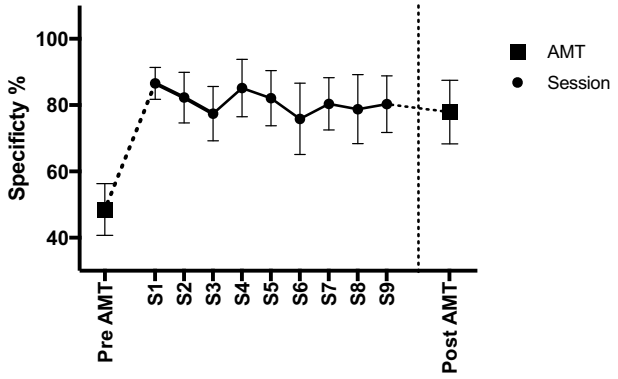


Fig. 3. Mean score with standard error on AMT for pre- and post-measurement, in between mean scores (and 95% Confidence Interval) for each online session.

3.3. Effects of c-MeST on depressive symptoms and related states (sadness, happiness, being tense when a painful memory arose) and processes (rumination, worry, thought suppression)

Visual analysis of depressive symptoms and related states (sadness, happiness, being tense when a painful memory arose) and processes (rumination, worry, thought suppression) per participant for 19 participants using the levels of each phase using mean scores and the trend of each phase using least squares regression do not suggest a clear effect. Visual Analysis of depressive symptoms (PHQ-9) of 19 participants is shown in Fig. 4.

Raw data and graphs of all other participants are available online (Martens et al., 2018). Combining p-values in a multiplicative way indicated that there was no significant impact of the training on depressive symptoms (PHQ, p = .11), rumination (RRS Brooding, p = .67), worrying about the future (single item, p = .68), worrying about the past (single item, p = .36), the combination of unwanted thoughts or images, being tense when a painful memory arises and thought suppression (three items combined, p = .33), sadness (single item, p = .65) or happiness (single item, p = .06). Test Statistics, which indicate the size of the effect, P values for each participant, and the multiplicative p-values are shown in Table 2.

Due to the lack of a clear hypothesis about where a clinically significant effect of c-MeST on secondary measures could have been expected, the training phase and follow-up phase were combined for the previously described analyses. Post-hoc visual analyses of repeated depressive symptomatology (PHQ-9) for the three separate phases (baseline versus training phase versus follow-up phase) per participant did not reveal an effect (according to the criteria defined by Kratochwill et al., 2010) that occurred during the training phase which disappeared during the follow-up neither, nor an effect that only appeared during follow-up. Moreover, analyses of the mean scores of the baselines of participants (n = 19) revealed that from a day to day basis 8 participants reported a mean of mild symptoms (< 5 on the PHQ-9; Kroenke et al., 2001) and 6 reported moderate depressive symptoms (5–9 on PHQ-9). With only 2 participants reporting moderately severe (10–14) or severe (> 15) depressive symptomatology, the margin for c-MeST to have caused a statistically significant effect was very limited. Post hoc visual analyses are available as Table A.3 (Appendix).

3.4. Feasibility – Training experiences

Results indicate that no clear difference was found for difficulty of words between sessions and that participants rated the cue words as average on difficulty (M = 4.9, SD = 1.94). For the feedback of the classifier, no difference was found between how correct participants experienced the feedback over sessions. On average they experienced the classifier as correct (M = 6.62, SD = 2.17). The length of the sessions was experienced on average as 'just right' and 'a bit too long' (M = 3.6, SD = 0.68). We assessed recurrent themes in participants' answers when asked about how they experienced the training and if they had any other comments. Most participants (n = 16) initially reported that c-MeST was difficult, frustrating or exhausting. Some noticed that they got better at retrieving memories more quickly (n = 8), some experienced the training as dull (n = 4), and some experienced less stress than in the face-to-face pre-intervention measurement of specificity (n = 3). Other comments of note were that one person was in doubt when to skip one cue and going to the next one and how long they should try to come up with a specific memory, and one person reported some distrust about sharing personal memories in an online platform.

4. Discussion

This study examined the impact of online Memory Specificity Training (c-MeST) on rAMS and on depressive symptoms and related

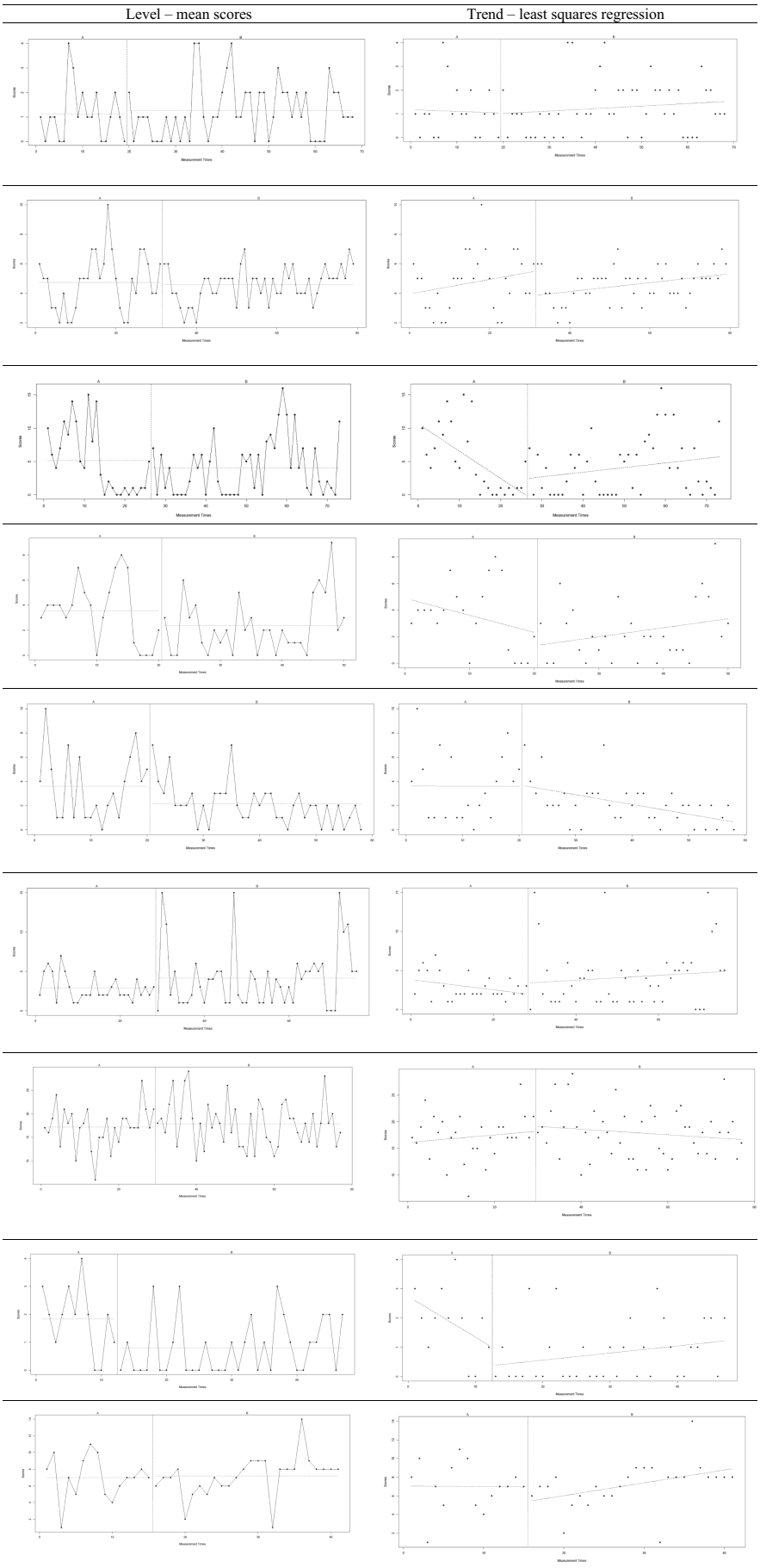


Fig. 4. Visual analysis of AB-design of depressive symptoms (PHQ-9) for 19 participants showing (1) level of each phase using mean scores and showing (2) the trend of each phase using least squares regression.

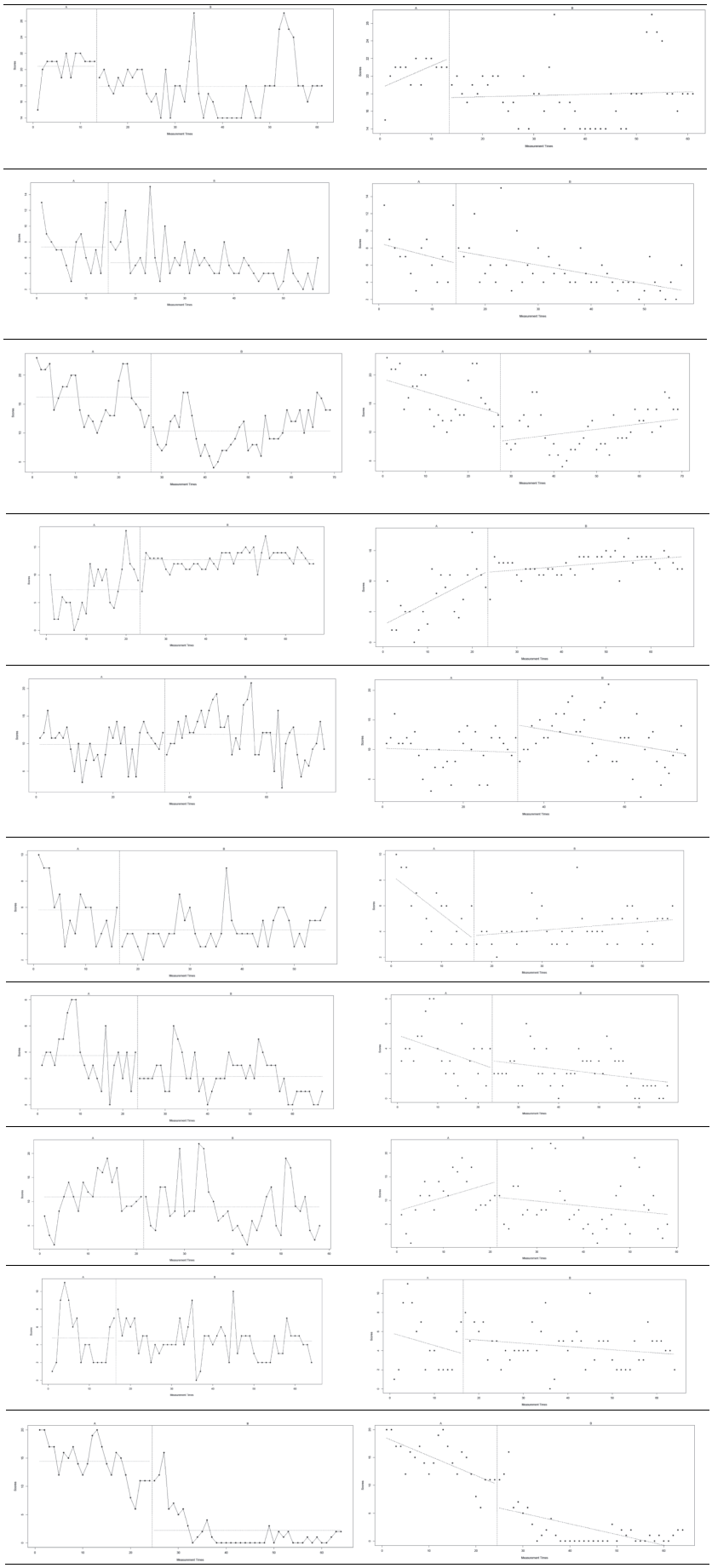


Fig. 4. (*continued*)

Table 2
Impact of online MeST on randomized baselines of PHQ-9, RRS Brooding, worry about the future, worry about the past, a combination of the three items inspired by the revised Impact of Events scale (unwanted thoughts or images, being tense when a painful memory arises, thought suppression), sadness and happiness: test statistic (TS), randomization test per participant (p-value) and an overall p-value (multiplicative meta-analysis).

Participant	PHQ-9		RRS Brooding		Worry future		Worry past		Unwanted thoughts, being tense, thought suppression		Sadness		Happiness	
	TS (A-B)	p	TS (A-B)	p	TS (A-B)	p	TS (A-B)	p	TS (A-B)	p	TS (A-B)	p	TS (B-A)	p
1	−0.07	0.40	−0.33	0.55	−0.15	0.45	−0.31	0.25	−1.20	0.85	−1.03	0.85	−0.09	0.70
2	0.14	0.25	0.85	0.30	2.08	0.85	0.65	0.10	2.55	0.10	1.65	0.10	−0.54	0.10
3	1.29	0.75	0.62	0.80	−0.49	0.75	0.02	0.75	−0.29	0.65	0.64	0.65	0.18	0.75
4	1.30	0.40	0.66	0.50	0.45	0.65	0.61	0.45	0.95	0.50	0.32	0.60	−0.53	0.85
5	1.82	0.35	1.10	0.30	1.22	0.35	0.95	0.25	7.14	0.35	0.25	0.35	−0.11	0.35
6	−1.01	0.90	−0.27	0.10	−0.73	0.55	0.41	0.70	−1.14	0.85	−0.36	0.95	−1.78	0.10
7	−0.90	0.25	−1.36	0.50	0.17	0.15	0.10	0.15	0.78	0.15	−0.03	0.15	0.30	0.25
8	0.97	0.15	1.54	0.40	−0.49	0.40	0.80	0.05	2.96	0.05	1.11	0.25	0.48	0.50
9	−0.25	0.15	1.19	0.30	0.48	0.30	−0.88	0.55	−0.26	0.30	−0.09	0.75	0.44	0.25
10	2.32	0.05	−4.44	1.00	0.24	1.00	0.26	1.00	−3.02	1.00	0.66	0.70	1.21	0.20
11	2.65	0.75	2.12	0.55	0.38	0.80	0.96	0.75	3.50	0.75	0.59	0.75	0.42	0.75
12	6.05	0.45	1.80	0.50	0.55	0.50	0.89	0.25	1.36	0.15	1.19	0.20	1.07	0.10
13	−5.54	0.50	−1.05	0.65	0.29	0.20	0.58	0.55	1.50	0.20	−1.83	0.50	0.52	0.35
14	−1.78	0.60	0.37	0.80	−1.00	0.60	1.17	0.20	−1.32	0.10	−0.84	0.55	−0.93	0.10
15	1.64	0.20	−0.24	1.00	−0.23	0.60	−0.24	0.60	−1.82	0.60	−0.59	0.60	−0.02	0.55
16	1.59	0.50	1.05	0.55	0.45	0.25	1.21	0.70	1.59	0.60	0.22	0.50	0.30	0.40
17	1.91	0.25	0.80	0.15	0.86	0.35	−0.07	0.60	0.73	0.40	0.33	0.45	1.18	0.30
18	/	/	/	/	/	/	/	/	/	/	/	/	/	/
19	0.47	0.75	0.85	0.70	−1.38	0.75	−0.75	0.75	−0.63	0.75	0.69	0.80	0.57	0.65
20	11.98	0.30	6.54	0.40	3.28	0.60	2.07	0.40	5.36	0.60	5.24	0.30	3.27	0.30
Meta-analysis		0.11		0.67		0.68		0.36		0.33		0.65		0.06

states (sadness, happiness, being tense when a painful memory arose) and processes (rumination, worry, thought suppression). As predicted, the hypothesis that rAMS can be remediated by online training was supported by significant differences between the pre-intervention scores and scores in the follow-up assessment, three months after training. This is in line with previous studies in which a group version was tested (Eigenhuis et al., 2017; Neshat-Doost et al., 2013; Raes et al., 2009b; Werner-Seidler et al., 2018). In-person, group MeST in most of these previous studies also includes psycho-education on memory problems in depression (session one), psycho-education and exercises on how to notice when one is thinking on an overgeneral level and how to tackle that (STOP-model, session four), and therapist-plus-group interaction. The results of the current study support the idea that mere memory specificity trials suffice to improve AMS, which is in line with the results of Takano et al. (2017b).

However, session-by-session scores of memory specificity indicate that participants have less trouble retrieving specific memories when done so with no time limit, at home, online, with words (neutral, positive, and negative affect) of an Age of Acquisition lower than six and in combination with exercises without cue words. Within Session 1 AMS seemed to have improved to saturation. Although one could argue that a mean score of 86.59% (SD = 10.26) still leaves room for improvement, this score is comparable to post-intervention measures of memory specificity in studies which used reduced memory specificity as an inclusion criterion; ranging from 59.5% (Takano et al., 2017b; improvement by 25.2%) to 66.67% (Werner-Seidler et al., 2018; improvement by 23.22%). A first potential explanation for this sudden increase is that participants are less stressed when retrieving memories at home, without a time limit. A second relevant difference between c-MeST and the AMT is that c-MeST contained trials with neutral cues and trials with exercises of the day. These two types of trials generated the highest scores (see Table 1) and thus have contributed to the higher scores on c-MeST sessions in comparison to the AMT which only included negative and positive cues. Theories regarding AMS suggest that reduced specificity can be driven by avoidance of negative emotions (Williams et al., 2007). We might therefore expect reduced specificity in a test involving negative cues compared to those involving neutral cues. The inclusion of neutral cues within c-MeST sessions may then have led to what appeared to be an improvement in specificity but which was actually a difference in specificity between tests involving cue words of different valence. A third potential explanation is that, although test-retest reliability has previously been assessed for AMTs without immediate feedback (for an overview, see Griffith et al., 2012), a practice effect may have due to the feedback of the AMT at intake plus the first session of c-MeST, and/or due to drawing participants' attention to the concept of specificity. Nevertheless, by the follow-up AMT assessment that included only positive and negative cues, specificity remained significantly higher than at baseline.

As a result of the saturation effect, it remains unclear how much training participants need to achieve a lasting effect at a face-to-face post-intervention assessment. As this is the first assessment of MeST using session-to-session scores, it raises questions about improvements in memory specificity between pre- and post-interventions measurements of in-person group MeST (Raes et al., 2009b; Werner-Seidler et al., 2018). It is possible that only one session of training is needed to improve scores on an AMT, but the design employed here precludes us from concluding whether this is the case. Unfortunately, it was not possible to check how much time participants spent in each session. An increase in speed in retrieving specific memories, for example, might be another important variable that improves through training and which may not be realized by the first session. Another useful addition to future studies examining the optimal dose would be the use of transfer tasks, to control for the overlap of the AMT and MeST, or to examine change in the neurobiological processes underlying reduced specificity from pre- to post- intervention (Barry et al., 2018).

The second hypothesis that c-MeST would lead to a decrease in depressive symptoms and related processes and symptoms were not supported. We used a multiple baseline across participants design, which can be regarded as adding information to results from previous studies using only pre- and post-measurements (Eigenhuis et al., 2017; Raes et al., 2009b) and follow-up (Neshat-Doost et al., 2013; Werner-Seidler et al., 2018). Visual and statistical analysis showed no

significant effect of c-MeST on depressive symptoms and related states (sadness, happiness, being tense when a painful memory arose) and processes (rumination, worry, thought suppression). As no statistically significant impact was found of c-MeST on secondary processes and symptoms, these results should be interpreted with caution and more research is needed to examine the relationship between c-MeST and symptomatology, particularly amongst groups of people who are currently depressed. It is of note that rAMS is evident amongst people with a history of depression who are not currently symptomatic but amongst whom their rAMS is predictive of subsequent increases in symptoms (Kleim and Ehlers, 2008). It may be the case that amongst at-risk samples without clinical diagnoses, the benefits of MeST are realized in terms of preventing further increases in symptom severity that might otherwise be expected amongst such people. Future research must examine if c-MeST shows an impact on symptoms in samples with higher levels of symptomatology than in this at-risk group, or examine the value of c-MeST in preventing increases in symptoms amongst at-risk groups using a longer follow-up than three months.

Nevertheless, the results of this study suggest that online remediation of rAMS is feasible. No participants dropped out during the training (although it is worth mentioning that one participant was tired of the specificity exercises and declined a post-intervention assessment) and the mean amount of trials made can be regarded as satisfying (M = 94,65%, SD = 8.179). Previous studies had drop-outs (e.g. 41.2% in Raes et al. (2009a, 2009b) and 15.6% in Eigenhuis et al. (2017)) due to reasons which c-MeST could be an answer for; being discharged from the hospital where they attended MeST, being in a too depressed state to attend group training, or family circumstances. With depression being the leading cause of disability worldwide (WHO, 2017), the development of a scalable and accessible intervention targeting a transdiagnostic factor can be regarded as promising. Murray et al. (2016) posed 13 research questions to evaluate the potential of Digital Health Interventions (DHI) as scalable tools. Before this study took place, one could positively answer four of those questions: (a) there is a clear health need that MeST intends to address, (b) there is a defined population that could benefit from MeST: i.e. people experiencing rAMS, (c) there seems to be a credible causal explanation for MeST to achieve the desired impact, and (d) the possibility of harms has been adequately considered. The results of this study offer preliminary answers for three other research questions regarding whether c-MeST would be able to reach the population that needs it whilst preserving its core components and doing so with minimal cost. First, the fact that participants found c-MeST acceptable and feasible, increases the chance that when c-MeST is offered to other at-risk samples it is likely to be experienced in a similar way. For this study design, participants were required to train intensively during a short period of time, which resulted in some participants reporting that session length was too long. Outside of research contexts, one can assume that uptake can even be improved if participants can train at their own pace in their own chosen dosage. In addition, adapting this online prototype to a more gamified and visually more attractive version, which might include psycho-education on why to train memory specificity, might increase the motivation of people experiencing rAMS. Second, regarding the key component, this study adds to our knowledge regarding MeST in suggesting that AMS might be improved using only specificity trials. In contrast with other more complex DHIs consisting of several components, online MeST only has one core component (with different kinds of cue words). Third, on the cost of this DHI and its cost impact on users and health systems, the current results indicate that providing online MeST using an online classifier can be delivered at a low cost as no face-to-face clinical interaction is required. However, for future dissemination on larger scale bigger investments would be needed to allow more simultaneous users. As cost-effectiveness for MeST is unclear as well, future research might examine and compare costs of both methods of delivery.

The main limitation of this study is that participants were not asked to stop other treatments that they may have been receiving. Given the working mechanism of increasing memory specificity, a possible explanation is that participants benefit from c-MeST due to interaction effects of increasing memory specificity and psychological treatments or medication. Future research could focus on exclusively training memory specificity, or on examining interaction effects between MeST and psychological treatments. Indeed, future investigations could explore the value of c-MeST as an add-on to other interventions. Another limitation is that the absence of a between-subjects control group makes it impossible to rule out regression to the mean. However, our findings are in agreement with many others which used control groups and which observed improvements in specificity across groups with differing levels of AMS at pre-intervention (Neshat-Doost et al., 2013; Raes et al., 2009b). Future studies examining c-MeST should include a control group to account for this. An additional limitation of the current study is that we used a verbal AMT with feedback, in contrast with studies in which test-retest reliability of the AMT was examined using versions of the AMT without feedback (i.e. Takano et al., 2017a). This makes it difficult to compare pre-intervention specificity scores with specificity scores of other studies. Another potential issue of this version is that participants possibly already started to learn from the feedback during the pre-intervention measurement. However, the present study is novel in that it provides the first examination of session-to-session changes in specificity across MeST. Future studies using a similar design would be able to examine the relation between dosage and improvements in symptomatology once an improvement is found. Finally, the current study could have benefited from diagnostic assessment.

Future research could also aim to replicate these findings and extending our current knowledge by exploring the other research questions on c-MeST as a DHI, for example: what strategies should be used to support tailoring the DHI (c-MeST) to participants over time (for example the gamification of MeST). These results also imply that future research on the context-dependency of measuring memory specificity (results improved drastically from the face to face and time-limited pre measurement to the first online session from home) can be useful to increase insight into the phenomenon of memory specificity.

5. Conclusions

Online Memory Specificity Training can effectively improve autobiographical memory specificity amongst people at risk of increases in depression symptoms. No change in related processes (e.g., rumination and negative future thinking) and the symptoms of depression was

found. Future studies with longer follow-up durations are needed to test the validity of c-MeST as a an intervention for preventing increases in depression symptoms or as a stand-alone treatment amongst people with clinical diagnoses.

Funding

This work was supported by KU Leuven Program Funding Grant PF/10/005.

Declaration of Competing Interest

Dr. Raes is one of the developers of the original in-group face-to-face MeST. Kris Martens en Drs Takano and Raes are the developers of the online, computerized MeST (c-MeST). Kris Martens and Dr. Raes additionally receive payments for training workshops and presentations related to MeST. We wish to confirm that there are no other known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Appendix A

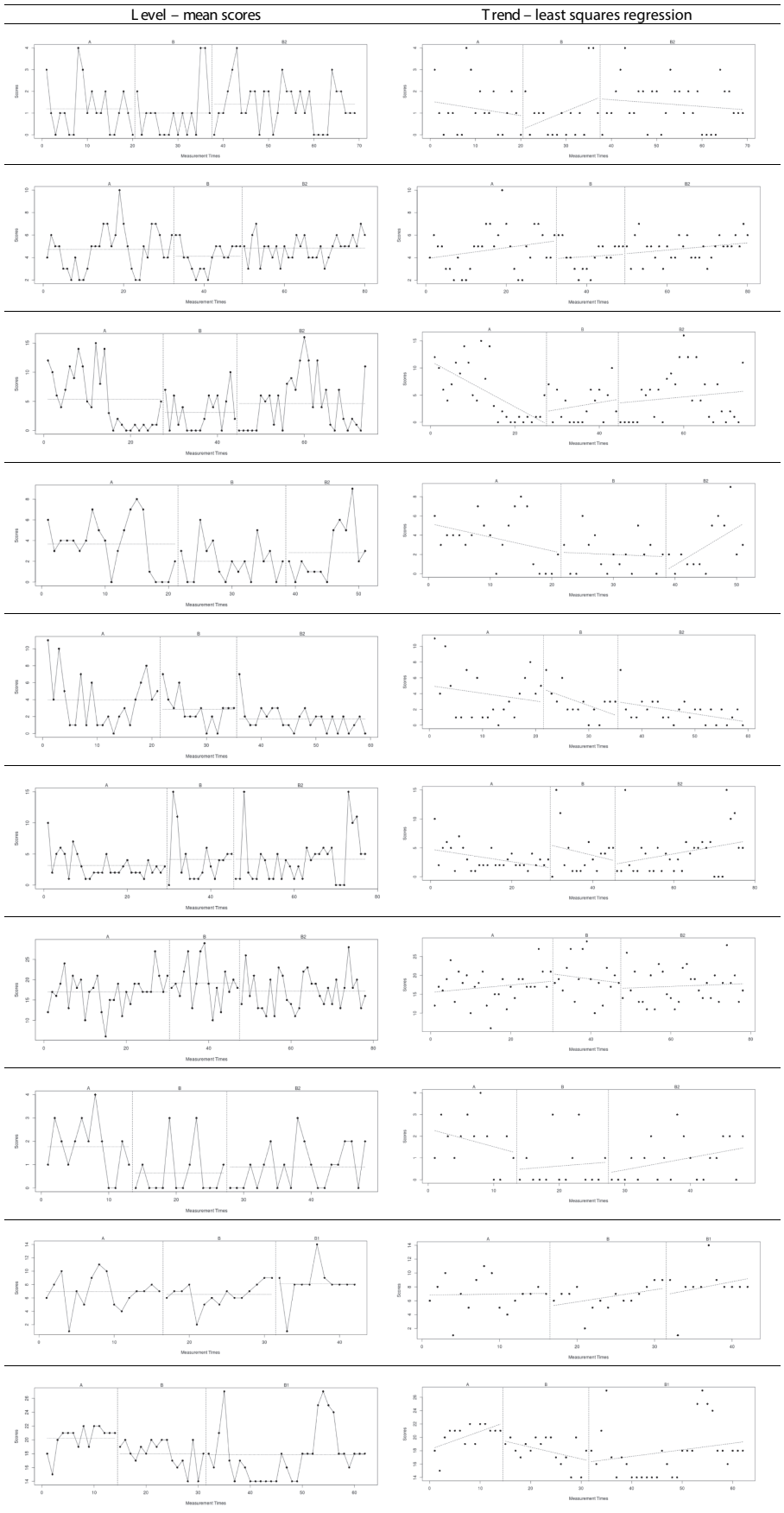
Table A.1
Cue words of AMT A, AMT B and nine c-MeST-sessions.

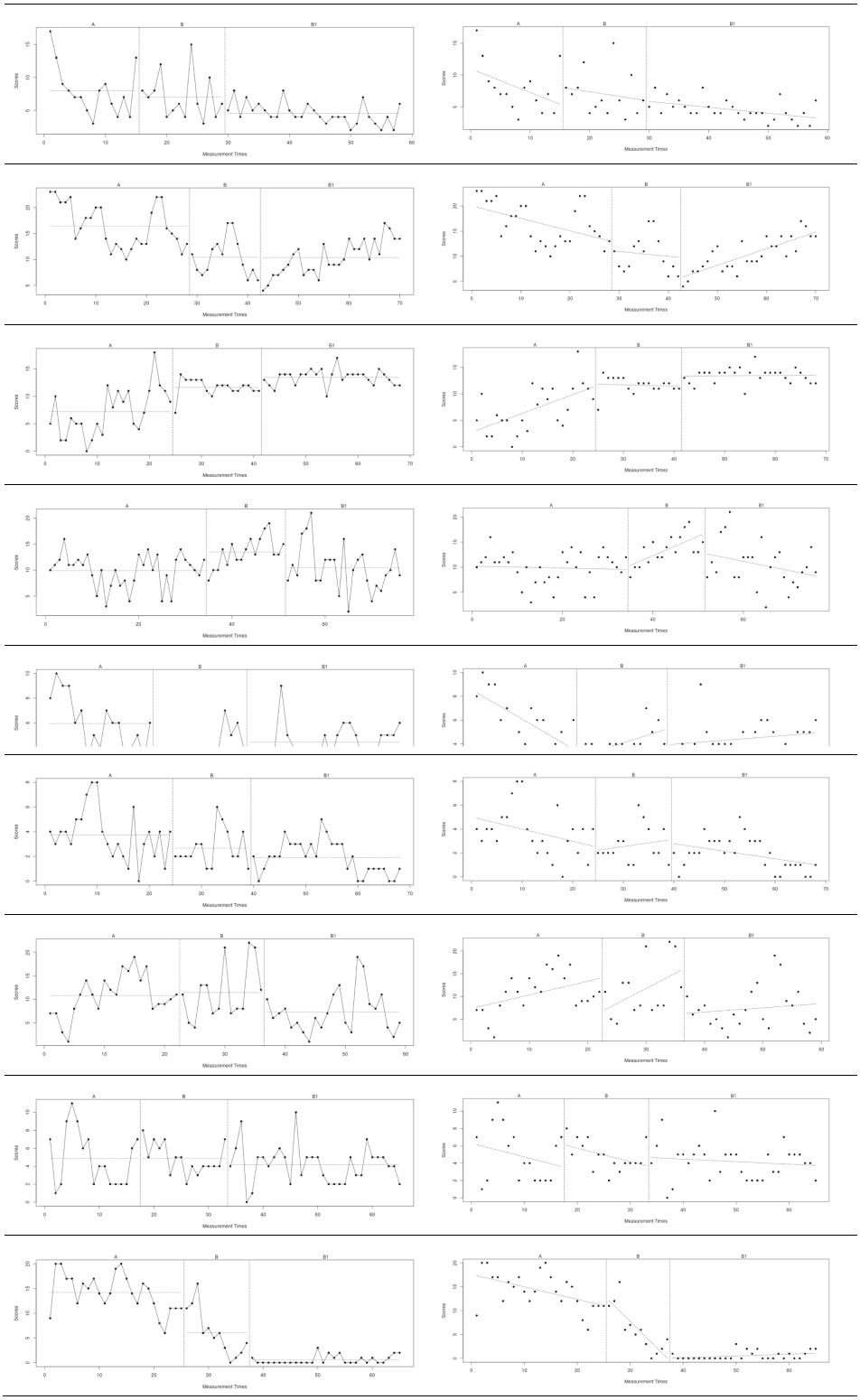
AMT A	AMT B	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9
Pleasant	Active	Pain	Accident	Captive	To lie	Ill	Rage	To stink	Quarrel	Odor
Mad	Furious	To cry	Angry	Dirty	Evil	False	Rot	Stupid	Sick	Ugly
Attentive	Interested	Broken	Weak	Wrong	Dirty	Lazy	Thick	Lost	Noise	Hunger
Hurt	Guilty	To work	Bank	Chair	Leg	Leaf	Table	Clock	Chin	Belly
Proud	Brave	Roof	Simple	Fish	Hot	Normal	Price	Letter	Hair	Tent
Angry	Helpless	Forest	Tree	Voice	Animal	News	Bicycle	Sauce	City	School
Social	Safe	In love	Holidays	Friend	Dear	Happy	Fun	Kiss	Merry	Feast
Clumsy	Sad	Handsome	Surprise	Free	Fun	Friendly	Smile	Funny	Laugh	Nice
Enthusiastic	Carefree	To kiss	Family	Smart	Well	Gift	Applause	Fine	Birthday	Music
Disillusioned	Anxious									

Table A.2
Specificity scores (%) per participant for pre and post measurement (Autobiographical Memory Test) and for each online session.

Participant	AMT Pre	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8	Session 9	AMT Post
1	50.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
2	60.00	100.00	90.91	100.00	100.00	100.00	100.00	/	100.00	100.00	100.00
3	40.00	90.91	63.64	63.64	81.82	63.64	54.55	81.82	72.73	72.73	50.00
4	60.00	100.00	100.00	81.82	100.00	81.82	100.00	63.64	90.91	100.00	90.00
5	40.00	81.82	90.91	90.91	100.00	100.00	81.82	90.91	100.00	90.91	100.00
6	10.00	81.82	72.73	81.82	72.73	81.82	63.64	100.00	72.73	72.73	70.00
7	50.00	90.91	63.64	45.45	72.73	72.73	90.91	72.73	90.00	72.73	70.00
8	40.00	81.82	63.64	/	100.00	90.91	72.73	81.82	100.00	90.91	50.00
9	60.00	81.82	72.73	72.73	/	90.91	54.55	72.73	/	72.73	90.00
10	60.00	63.64	36.36	45.45	54.55	36.36	9.09	36.36	45.45	45.45	/
11	50.00	72.73	90.91	63.64	90.91	63.64	90.91	72.73	90.00	81.82	90.00
12	0.00	90.91	90.91	90.91	100.00	81.82	81.82	90.91	72.73	81.82	50.00
13	60.00	81.82	81.82	72.73	72.73	63.64	72.73	63.64	81.82	72.73	80.00
14	50.00	90.91	90.91	72.73	90.91	100.00	100.00	72.73	81.82	100.00	50.00
15	60.00	90.91	81.82	66.67	81.82	90.91	72.73	66.67	81.82	90.91	100.00
16	50.00	100.00	90.91	90.91	100.00	90.91	90.91	100.00	90.91	/	90.00
17	60.00	80.00	81.82	72.73	100.00	87.50	90.91	90.00	18.18	/	80.00
18	50.00	70.00	100.00	/	36.36	54.55	44.44	88.89	44.44	45.45	50.00
19	60.00	90.91	100.00	90.91	90.91	100.00	81.82	100.00	81.82	63.64	100.00
20	60.00	90.91	81.82	90.91	72.73	90.91	63.64	81.82	81.82	90.91	70.00

Table A.3
Visual analysis of ABB'-design of depressive symptoms (PHQ-9) for 19 participants showing (1) level of each phase (baseline – training – follow-up) using mean scores and showing (2) the trend of each phase using least squares regression.





References

Barry, T.J., Chiu, C.P.Y., Raes, F., Ricarte, J., Lau, H., 2018. The neurobiology of reduced autobiographical memory specificity. Trends Cogn. Sci. 22, 1038–1049. https://doi.org/10.1016/j.tics.2018.09.001.

Barry, T.J., Sze, W.Y., Raes, F., 2019. A meta-analysis and systematic review of Memory Specificity Training (MeST) in the treatment of emotional disorders. Behav. Res. Ther. 116, 36–51. https://doi.org/10.1016/j.brat.2019.02.001.

Beaman, A., Pushkar, D., Etezadi, S., Bye, D., Conway, M., 2007. Autobiographical memory specificity predicts social problem-solving ability in old and young adults. Q. J. Exp. Psychol. 60, 1275–1288. https://doi.org/10.1080/17470210600943450.

Bomba, M., Marfone, M., Brivio, E., Oggiano, S., Broggi, F., Neri, F., Nacinovich, R., 2014. Autobiographical memory in adolescent girls with anorexia nervosa. Eur. Eat. Disord. Rev. 22, 479–486. https://doi.org/10.1002/erv.2321.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., Storms, G., 2014. Norms of age

of acquisition and concreteness for 30,000 Dutch words. Acta Psychol. 150, 80–84. https://doi.org/10.1016/j.actpsy.2014.04.010.
Bulté, I., Onghena, P., 2011. When the truth hits you between the eyes. A software tool for the visual analysis of single-case experimental data. *Methodology* 8, 104–114. https://doi.org/10.1027/1614-2241/a000042.
Bulté, I., & Onghena, P. (2013). The single-case data analysis package: Analysing single-case experiments with R software. *Journal of Modern Applied Statistical Methods*, *12*, 450–478. doi:10.22237/jmasm/1383280020.
Burcusa, S.L., Iacono, W.G., 2007. Risk for recurrence in depressoin. Clin. Psychol. Rev. 8, 959–985. https://doi.org/10.1016/j.cpr.2007.02.005.
Conway, M.A., Pleydell-Pearce, C.W., 2000. The construction of autobiographical memories in the self-memory system. Psychol. Rev. 107, 261–288. https://doi.org/10.1037//0033-295X.
Craske, M.G., 2018. Honoring the past, envisioning the future: ABCT's 50th anniversary presidential address. Behav. Ther. 49, 151–164. https://doi.org/10.1016/j.beth.2017.05.003.
Dalgleish, T., Williams, J.M.G., Golden, A.-M.J., Perkins, N., Barrett, L.F., Barnard, P.J., ... Watkins, E., 2007. Reduced specificity of autobiographical memory and depression: the role of executive control. J. Exp. Psychol. Gen. 136, 23–42. https://doi.org/10.1037/0096-3445.136.1.23.
Debeer, E., Raes, F., Claes, S., Vrieze, E., Williams, J.M.G., Hermans, D., 2012. Relationship between cognitive avoidant coping and changes in overgeneral autobiographical memory retrieval following an acute stressor. J. Behav. Ther. Exp. Psychiatry 43, S37–S42. https://doi.org/10.1016/j.jbtep.2011.04.002.
Eigenhuis, E., Seldenrijk, A., Schaik, A. V, Raes, F., & Oppen, P. V. (2017). Feasibility and effectiveness of Memory Specificity Training in depressed outpatients: a pilot study. Clin. Psychol. Psychother., 24, 269–277. doi:https://doi.org/10.1002/cpp.1995.
Goddard, L., Dritschel, B., Burton, A., 1996. Role of autobiographical memory in social problem solving and depression. J. Abnorm. Psychol. 105, 609–616. https://doi.org/10.1037/0021-843X.105.4.609.
Griffith, J.W., Sumner, J.A., Raes, F., Barnhofer, T., Debeer, E., Hermans, D., 2012. Current psychometric and methodological issues in the measurement of overgeneral autobiographical memory. J. Behav. Ther. Exp. Psychiatry 43, 21–31. https://doi.org/10.1016/j.jbtep.2011.05.008.
Hayes, S.C., 1981. Single case experimental design and empirical clinical practice. J. Consult. 49, 193–211.
Heyvaert, M., Onghena, P., 2014. Randomization tests for single-case experiments: state of the art, state of the science, and state of the application. J. Contextual Behav. Sci. 3, 51–64. https://doi.org/10.1016/j.jcbs.2013.10.002.
Heyvaert, M., Moeyaert, M., Verkempynck, P., Van den Noortgate, W., Vervloet, M., Ugille, M., Onghena, P., 2017. Testing the intervention effect in single-case experiments: a Monte Carlo simulation study. J. Exp. Educ. 85, 175–196. https://doi.org/10.1080/00220973.2015.1123667.
Hitchcock, C., Werner-Seidler, A., Blackwell, S.E., Dalgleish, T., 2017. Autobiographical episodic memory-based training for the treatment of mood, anxiety and stress-related disorders: a systematic review and meta-analysis. Clin. Psychol. Rev. 52, 92–107. https://doi.org/10.1016/j.cpr.2016.12.003.
Huber, J., Salatsch, C., Ingenerf, K., Schmid, C., Maatouk, I., Weisbrod, M., ... Nikendei, C., 2015. Characteristics of disorder-related autobiographical memory in acute anorexia nervosa patients. Eur. Eat. Disord. Rev. 23, 379–389. https://doi.org/10.1002/erv.2379.
Jing, H.G., Madore, K.P., Schacter, D.L., 2016. Worrying about the future: an episodic specificity induction impacts problem solving, reappraisal, and well-being. J. Exp. Psychol. Gen. 145, 402–418.
Kleim, B., Ehlers, A., 2008. Reduced autobiographical memory specificity predicts depression and posttraumatic stress disorder after recent trauma. J. Consult. Clin. Psychol. 76, 231–242. https://doi.org/10.1037/0022-006X.76.2.231.
Kratochwill, T.R., Hitchcock, J., Horner, R.H., Levin, J.R., Odom, S.L., Rindskopf, D.M., Shadish, W.R., 2010. Single-case design technical documentation. What Works Clearing House 1–34. https://doi.org/10.1037/e578392011-004.
Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.
Kroenke, K., Spitzer, R.L., Williams, J.B.W., Löwe, B., 2010. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen. Hosp. Psychiatry 32, 345–359. https://doi.org/10.1016/j.genhosppsych.2010.03.006.
Mackinger, H.F., Pachinger, M.M., Leibetseder, M.M., Fartacek, R.R., 2000. Autobiographical memories in women remitted from major depression. J. Abnorm. Psychol. 109, 331–334. https://doi.org/10.1037//0021-843X.109.2.331.
Martens, K., Barry, T. J., Takano, K., Onghena, P., & Raes, F. (2018). Efficacy of online Memory Specificity Training in patients with a history of depression. doi:10.17605/OSF.IO/GE6VS.
McNally, R.J., Lasko, N.B., Macklin, M.L., Pitman, R.K., 1995. Autobiographical memory disturbance in combat-related posttraumatic stress disorder. Behav. Res. Ther. 33, 619–630. https://doi.org/10.1016/0005-7967(95)00007-K.
Michiels, B., Heyvaert, M., Onghena, P., 2018. The conditional power of randomization tests for single-case effect sizes in designs with randomized treatment order: a Monte Carlo simulation study. Behav. Res. Methods 50, 557–575. https://doi.org/10.3758/s13428-017-0885-7.
Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A.-L., ... Brysbaert, M., 2013. Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. Behav. Res. Methods 45, 169–177. https://doi.org/10.3758/s13428-012-0243-8.
Murray, E., Hekler, E.B., Andersson, G., Collins, L.M., Doherty, A., Hollis, C., ... Wyatt, J.C., 2016. Evaluating digital health interventions: key questions and approaches. Am. J. Prev. Med. 51, 843–851. https://doi.org/10.1016/j.amepre.2016.06.008.
Neshat-Doost, H.T., Dalgleish, T., Yule, W., Kalantari, M., Ahmadi, S.J., Dyregrov, A., Jobson, L., 2013. Enhancing autobiographical memory specificity through cognitive training. Clin. Psychol. Sci. 1, 84–92. https://doi.org/10.1177/2167702612454613.
Nolen-Hoeksema, Susan. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology*, *109*, 504–511. doi:10.1037.0021)843X.109.3.504.
Nolen-Hoeksema, S., Morrow, J., 1991. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta Earthquake. J. Pers. Soc. Psychol. 61, 115–121. https://doi.org/10.1037/0022-3514.61.1.115.
Onghena, P., Edgington, E.S., 2005. Customization of pain treatments - single-case design and analysis. Clin. J. Pain 21, 56–68.
Park, R.J., Goodyer, I.M., Teasdale, J.D., 2002. Categoric overgeneral autobiographical memory in adolescents with major depressive disorder. Psychol. Med. 32, 267–276. https://doi.org/10.1017/S0033291701005189.
Raes, F., Hermans, D., Williams, J.M.G., Demyttenaere, K., Sabbe, B., Pieters, G., Eelen, P., 2005. Reduced specificity of autobiographical memory: a mediator between rumination and ineffective social problem-solving in major depression? J. Affect. Disord. 87, 331–335. https://doi.org/10.1016/j.jad.2005.05.004.
Raes, F., Schoofs, H., Hoes, D., Hermans, D., Van Den Eede, F., Frank, E., 2009a. 'Reflection' en 'brooding' als subtypes van rumineren: een herziening van de Ruminative Response Scale. Gedragstherapie 42, 205–214.
Raes, F., Williams, J.M.G., Hermans, D., 2009b. Reducing cognitive vulnerability to depression: a preliminary investigation of MEmory Specificity Training (MEST) in inpatients with depressive symptomatology. J. Behav. Ther. Exp. Psychiatry 40, 24–38. https://doi.org/10.1016/j.jbtep.2008.03.001.
Starr, L.R., Davila, J., 2012. Responding to anxiety with rumination and hopelessness: mechanism of anxiety-depression symptom co-occurrence? Cogn. Ther. Res. 36, 321–337. https://doi.org/10.1007/s10608-011-9363-1.
Sumner, J.A., Griffith, J.W., Mineka, S., 2010. Overgeneral autobiographical memory as a predictor of the course of depression: a meta-analysis. Behav. Res. Ther. 48, 614–625. https://doi.org/10.1016/j.brat.2010.03.013.
Sutherland, K., Bryant, R.A., 2008. Social problem solving and autobiographical memory in posttraumatic stress disorder. Behav. Res. Ther. 46, 154–161. https://doi.org/10.1016/j.brat.2007.10.005.
Takano, K., Mori, M., Nishiguchi, Y., Moriya, J., Raes, F., 2017a. Psychometric properties of the written version of the autobiographical memory test in a Japanese community sample. Psychiatry Res. 248, 56–63. https://doi.org/10.1016/j.psychres.2016.12.019.
Takano, K., Moriya, J., Raes, F., 2017b. Lost in distractors: reduced autobiographical memory specificity and dispersed activation spreading over distractors in working memory. Behav. Res. Ther. 94, 19–35. https://doi.org/10.1016/j.brat.2017.04.005.
Takano, K., Ueno, M., Moriya, J., Mori, M., 2017c. Unraveling the linguistic nature of specific autobiographical memories using a computerized classification algorithm. Behav. Res. Methods 49, 835–852. https://doi.org/10.3758/s13428-016-0753-x.
Takano, K., Gutenbrunner, C., Martens, K., Salmon, K., Raes, F., 2018. Computerized scoring algorithms for the autobiographical memory test. Psychol. Assess. 30, 259–273. https://doi.org/10.1037/pas0000472.
Treynor, W., Gonzalez, R., Nolen-Hoeksema, S., 2003. Rumination reconsidered: a psychometric analysis. Cogn. Ther. Res. 27, 247–259. https://doi.org/10.1023/A:1023910315561.
Tzemou, E., Birchwood, M., 2007. A prospective study of dysfunctional thinking and the regulation of negative intrusive memories in bipolar 1 disorder: implications for affect regulation theory. Psychol. Med. 37, 689–698. https://doi.org/10.1017/S0033291706009470.
Van der Does, A.J.W., Barnhofer, T., Williams, J.M.G., 2003. The Major Depression Questionnaire (MDQ).
Watkins, E.R., Baeyens, C.B., Read, R., 2009. Concreteness training reduces dysphoria: proof-of-principle for repeated cognitive bias modification in depression. J. Abnorm. Psychol. 118, 55–64. https://doi.org/10.1037/a0013642.
Weiss, D. S., & Marmar, C. R. (1995). Impact of events scale - revised. *Assessing Psychcological Trauma and PTSD*: *A Practitioner's Handbook N.Y*: *Guildford*, *1995*.
Werner-Seidler, A., Hitchcock, C., Bevan, A., McKinnon, A., Gillard, J., Dahm, T., ... Dalgleish, T., 2018. A cluster randomized controlled platform trial comparing group MEmory Specificity Training (MEST) to group psychoeducation and supportive counselling (PSC) in the treatment of recurrent depression. Behav. Res. Ther. 105, 1–9. https://doi.org/10.1016/j.brat.2018.03.004.
Williams, J.M.G., 2007. Capture and rumination, functional avoidance, and executive control (CaRFAX): three processes that underlie overgeneral memory. Cognit. Emot. 20, 548–568. https://doi.org/10.1080/02699930500450465.
Williams, J.M.G., Broadbent, K., 1986. Autobiographical memory in suicide attempters. J. Abnorm. Psychol. 95, 144–149. https://doi.org/10.1037/0021-843X.95.2.144.
Williams, J.M.G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., Dalgleish, T., 2007. Autobiographical memory specificity and emotional disorder. Psychol. Bull. 133, 122–148. https://doi.org/10.1037/0033-2909.133.1.122.
Wilson, F.C.L., Gregory, J.D., 2017. Overgeneral autobiographical memory and depression in older adult: a systematic review. Aging Ment. Health 22. https://doi.org/10.1080/13607863.2017.1326461.
World Health Organization, 2017. Depression and Other Common Mental Disorders: Global Health Estimates. *World Health Organization*, pp. 1–24 (https://doi.org/CC BY-NC-SA 3.0 IGO).