Journal of Affective Disorders 244 (2019) 187–195

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Research paper

Imagery-based treatment for test anxiety: A multiple-baseline open trial

Jessica N. Prinz[a,*], Eran Bar-Kalifa[b], Eshkol Rafaeli[c], Haran Sened[c], Wolfgang Lutz[a]

[a] Department of Clinical Psychology and Psychotherapy, University Trier, Trier, Germany
[b] The Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[c] Department of Psychology and the Gonda Brain Science Center, Bar-Ilan University, Ramat Gan, Israel

ARTICLE INFO

Keywords:
Imagery rescripting
Test anxiety
CBT
Schema therapy

Background: Many students are affected by test anxiety (TA), which involves considerable distress and can lead to reduced performance. Despite its prevalence, TA has been under-studied; specifically, few effective and brief interventions have been identified to date. In our work, we explore the adaptation of imagery, an emotion-focused treatment technique used widely in the treatment of other anxiety disorders, for addressing TA.
Methods: In a two-site concurrent multiple-baseline pilot study (n = 31), we examined the effectiveness of a targeted six-session protocol developed for the treatment of TA, which integrates traditional cognitive behavioral techniques with imagery work.
Results: The protocol was well-accepted by clients. We found that students' test anxiety levels did not drop between the recruitment and pre-intervention assessments, but did drop significantly from recruitment or baseline to the delayed follow-up (Cohen's d = 0.75 and 0.84, respectively). We also found evidence for session-level processes tying the quality of the imagery work with session efficacy.
Limitations: The study involved a relatively small sample size, leading to weaker power to detect treatment effects. Moreover, some clients did not have any exams scheduled before the delayed follow-up assessment. Due to the scheduling intensity of the sessions, some clients had insufficient time to implement or practice skills and to complete tasks discussed in the sessions.
Conclusions: This study provides preliminary evidence for the utility of integrating imagery work with traditional cognitive-behavioral techniques for treating test anxiety.

1. Introduction

Test anxiety (TA), a condition which affects as many as 20% of students (see Hill and Wigfield, 1984; Holm-Hadulla et al., 2009), involves excessive negative thoughts about poor performance or failure in exams or other situations characterized by evaluation, as well as about the possible consequences of such performance. Students with high TA are more likely to be female, and have perfectionistic tendencies, which are maladaptive or unhelpful (Eum and Rice, 2011). TA is manifested in phenomenological, physiological, and behavioral responses (Zeidner, 1998; Zeidner and Matthews, 2010). Phenomenologically, it often involves intense fear, and can be accompanied by panic attacks, feelings of hopelessness, inferiority, and desperation (e.g., Fehm and Fydrich, 2011). Physiologically, TA involves a stress response which includes hyper-activation of the hypothalamic-pituitary-adrenal axis (e.g., Morris and Liebert, 1970). Behaviorally, TA often culminates in sleep disturbances, loss of appetite, and procrastination before the exam, and in disrupted concentration, sweating, shaking, and blackouts during the exam (for review, see Fehm and Fydrich, 2011, Zeidner, 2014).

There is considerable and rising demand for psychological interventions to address TA. For example, among the students receiving services at one large German university's mental health center, 56% were treated for TA; this figure involved a 20% increase compared to that of 10 years prior (Holm-Hadulla et al., 2009).

Several approaches have been developed for the treatment of TA, including skill-focused techniques, cognitive and behavioral interventions, Gestalt techniques, meditation, and physical exercise. A meta-analysis examining the effectiveness of these approaches (Ergene, 2003) reported that programs which used behavioral or cognitive techniques, as well as ones that combined these techniques with a skill-focused approach, produced the largest effects (0.80–1.22). Gestalt therapy, meditation, and physical exercises showed small effect sizes (0.15). In addition, combined group + individual therapy showed greater efficacy than group or individual programs (Ergene, 2003). Despite their promise, existing approaches for the treatment of TA have

* Corresponding author.
E-mail address: prinzj@uni-trier.de (J.N. Prinz).

https://doi.org/10.1016/j.jad.2018.10.091
Received 6 March 2018; Received in revised form 4 September 2018; Accepted 5 October 2018
Available online 06 October 2018
0165-0327/

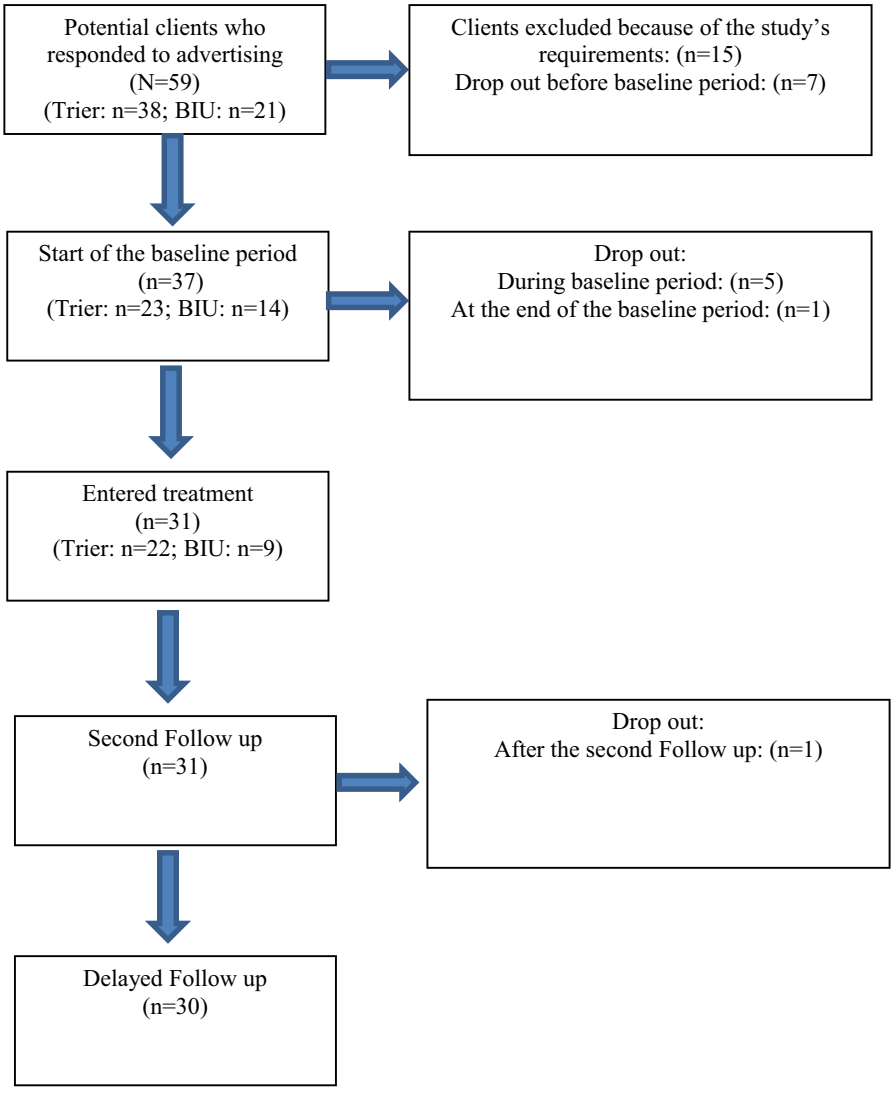

Fig. 1. Recruitment and retention flow chart.

room for improvement. In particular, several authors have noted the need to go beyond traditional cognitive techniques in treating anxiety conditions (TA: Brown et al., 2011; social anxiety: Wild and Clark, 2014). As Brown and her colleagues note, standard cognitive tools such as cognitive restructuring may be too time-consuming to implement during time-pressured evaluative situations such as tests. Additionally, as Wild and Clark suggest, surface level negative thoughts (e.g. "I'm blushing—they must think I am odd") often result from deeper underlying negative emotional beliefs (e.g., "I am defective, and that defectiveness is easily seen unless I hide it") and from even deeper schemas (e.g., defectiveness and shame); it may prove more effective to address these schemas and underlying negative emotional beliefs themselves rather than combat the automatic thoughts.

One promising approach for addressing underlying emotional beliefs involves the emotion-focused technique of imagery. As recent reviews (e.g., Ji et al., 2016; Holmes and Mathews, 2010) have noted, mental imagery gets to the core of individuals' distress in a very direct way. In particular, images of aversive experiences activate memories and associated emotions (Holmes and Mathews, 2010) and do so more intensively than verbal prompts (Holmes and Mathews, 2005). As a consequence, imagery work provides a more powerful means of accessing the distress and of intervening with it than do purely verbal techniques (for review, see Holmes et al., 2007; Rafaeli et al., 2016; Rafaeli et al., 2014).

Imagery work's effectiveness may be due, at least in part, to the fact that images activate early emotional memories which are encoded non-verbally, and thus provide access to the very situations in which negative (and sometimes traumatic) associations were established (e.g., Arntz, 2014). Gaining access to the memory of these situations can be helpful, as it can clarify what underlies the intense emotional reaction that still plagues the client, even in the absence of present-day danger. Prior research has shown the effect of rescripting early memories in imagery within the treatment of social anxiety disorder (Nilsson et al., 2012). Similarly, Reiss et al. (2017) found that a group intervention combining CBT and Imagery Rescripting (ImRs) as treatment for students with TA led to a significant reduction in symptoms, with the main benefits ascribed to the ImRs component. These results show that the addition of ImRs alone improves outcome compared to stand-alone CBT treatment. We can assume that the potency of the ImRs work—i.e., how accessible and changeable the emotional experience of the client becomes - will be associated with outcome changes.

Indeed, revising or changing negative images can bring about emotional or symptomatic relief. At times, this disruption could be focused on past events (and on changing their meaning); at other times, the disruption can be of present-day or even future-time images. In addition, positive imagery can be used to access adaptive emotional states (Ji et al., 2016). With these ideas in mind, the current pilot study was designed to examine the effectiveness of a new protocol integrating

CBT and imagery-based techniques into a brief intervention for TA. With clients recruited at two sites (Trier University in Germany, and Bar-Ilan University in Israel), we implemented a multiple-baseline study. We chose to use a multiple-baseline open trial design for several reasons. First, we wanted to ensure that throughout varying lengths of baseline-assessment periods, which also differed in the number of baseline assessments, we will see no decline in TA symptoms prior to the commencement of the treatment itself. Second, the symptomatology of TA may be strongly time related i.e., increase as exams are nearer. There is a large benefit to having multiple time lags as it can show whether a significant change occurs during the pre-treatment assessments and during the post-treatment assessments. . Clients' TA was assessed for 2–6 weeks prior to treatment; they then received 6 sessions over the course of 3 weeks, and were followed up at both 1 week and 3–7 weeks after the completion of the protocol. Additionally, symptom measures (TA) and process measures (an emotional process item, as well as session evaluation scales) were collected during each of the treatment sessions. These data allowed us to test the following hypotheses: (1) *Acceptability:* Clients will find the augmented protocol (integrating CBT and imagery-work techniques) acceptable. (2) *Outcome:* Clients treated with the protocol will show reduced test anxiety following treatment compared to pre-treatment. (3) *Process:* The potency of imagery work will be associated with session-level outcomes (i.e., improved working alliance and session quality), and with subsequent reductions in test anxiety.

2. Method

2.1. Clients: recruitment and demographics

Fig. 1 presents a flow chart detailing client recruitment and retention. In the Trier site, clients were recruited using flyers at the campus health service, ads placed in the newsletters of two local universities, as well as a brief invitation given to participants in a test anxiety prevention workshop initiated by the student-union at the University of Trier. In the BIU site, clients were recruited using flyers posted throughout campus and describing the IRB-approved study. In both sites, participation in the treatment was at no cost. To be eligible for the study, clients had to meet the following criteria: Test Anxiety Inventory (TAI; Spielberger, 1980) scores > 54 (one standard deviation above the average TAI score), no suicidality (a value 1 or less in the BDI-II suicide item), and no current therapy addressing test anxiety. The sample size was determined based on the availability of therapists. Fifty-nine potential clients (38 at Trier, 21 at BIU) were screened for eligibility. Fourteen were excluded because of their TAI scores and one had no exams at the end of her study period. Seven additional potential clients dropped out before the baseline period, most often because of time concerns. Thirty-seven (24 at Trier, 14 at BIU) met the inclusion requirements. Of these, five clients dropped out during the baseline period, and one at the end of the baseline period. Thus, 31 clients (23 females) started treatment (22 at Trier, 9 at BIU), and all completed the entire 6-session protocol. One client, from BIU, dropped out after the first follow-up assessment; her data were used in all analyses (using carry-forward to complete the missing assessment).

The 31 clients ranged in age from 19 to 53 years ($M = 25.9$, $SD = 6.3$). At Trier, twenty-one of the clients were university students, and one was a trainee in a nursing program. At BIU, all clients were students. At both sites, students came from diverse fields of study, most commonly computer science, law, business studies, and psychology, with a median of 4 years of post-secondary study at Trier ($SD = 3.7$, min = 1, max = 16) and a median of 2 years of post-secondary study at BIU ($SD = 1.1$, min = 1, max = 4. Informed consent was obtained from all clients before the treatment started.

2.2. Therapists: training and supervision

In the Trier site, four of the six therapists were graduates of a masters' program in psychology, and were awaiting the beginning of their clinical traineeship; the remaining two therapists were PhD students in clinical psychology, each with prior experience of one year as a clinician. In the BIU site, the therapists were four PhD students in clinical psychology, each with prior experience of two years as a clinician. All therapists were trained in the treatment protocol in a two-day workshop which involved modeling and role-plays of the intervention modules. In addition, the therapists reviewed their cases in a weekly supervised group, guided by an experienced psychotherapist.

2.3. Treatment protocol

A short-term imagery-based treatment protocol for TA was developed (for the full protocol, see www.osf.io/hraqd). The protocol involves six sessions lasting approximately 50 minutes each, carried out over three weeks. The sessions integrated traditional cognitive behavioral and imagery techniques; each was also followed by some homework assignment, aimed at practicing the contents of the session and/or preparing for the next one. Table 1 reviews the key cognitive-behavioral and imagery-related tasks of the sessions.

The cognitive behavioral components of the sessions were based on Safran et al. (2005) protocol. The imagery components of the sessions were based on ideas from Wild and Clark (2014) as well as Arntz and Weertman (1999). Imagery was guided by the therapist and conducted out loud so that the client's experiences became accessible to the therapist. This allowed the therapist to present questions or suggestions when needed, and thus helped ensure that the imagery work would be as emotionally compelling as possible.

Clients were encouraged to undertake imagery with their eyes shut. All clients did so willingly. Therapists were instructed to use a brief body scan – a mental scan of the major muscle groups from head to toe - to help clients relax and ease into the imagery work. This was always done in session 1, and was optional in subsequent sessions. Therapists then proceeded according to the following emphases in each session:

a) Session 1 was aimed at introducing the client to the protocol, providing psychoeducation regarding TA, and beginning to socialize the client to the use of imagery in a non-threatening way. The imagery component in this session involved safe place imagery only. Clients were asked to imagine a personal safe place in which they felt calm and relaxed. During the dialogue with the therapist, they explored their safe place, focusing on all four phenomenological elements: body sensations, emotions, cognitions, and behavioral tendencies. The post-session homework was to practice this safe place imagery (which clients were encouraged to continue doing for the duration of the therapy and beyond).
b) Session 2 was aimed at increasing the understanding of TA and its phenomenological elements, and at helping prepare the client to be able to monitor these. The imagery component in this session involved exploration of a situation relevant to TA. Clients were asked to imagine a distressing situation tied to exam or study situations, and again focus on all four phenomenological elements. The purpose

Table 1
Key aspects of the imagery-based treatment protocol.

Session	CB work	Imagery work
#1	Psychoeducation	Safe-place
#2	Automatic cognitions & behaviors	Exploration
#3	Alternative cognitions & behaviors	ImRS-past experience
#4	Learning skills	ImRS-past experience
#5	Test-taking skills	ImRS-future experience
#6	Consolidation & content review	ImRS-future experience

was to access the fear structure related to test situations in a way that bypasses routine verbal reports (for review, see Ji et al., 2016). The images that were evoked set the stage for a discussion about the client's unique phenomenology of TA. The post-session homework was to begin monitoring incidents of TA and to pay particular attention to their unique phenomenology.

c) Session 3 was aimed at utilizing the client's growing awareness of their own phenomenology in specific TA situations and began introducing the possibility of alternative behaviors and cognitions that could alter this phenomenology. The imagery component introduced the possibility of rescripting (past) distressing test- or studying-related situations (i.e., imagery with rescripting; see Arntz, 2014). The post-session homework involved practice in monitoring situations and identifying alternative behaviors and cognitions. It also involved completing a questionnaire about study skills.

d) Session 4 was aimed at providing additional practice with imagery and rescripting and at addressing study skill deficits. The imagery component was identical to that used in session 3 (i.e., imagery with rescripting). It was followed by a review of the client's study skills, with the aim of identifying ones that could be improved. The post-session homework involved practice in monitoring study-related situations, again identifying alternative behaviors and cognitions. It also involved completing a questionnaire about modes of learning and about test-taking. Finally, a specific behavioral task was also chosen as a homework goal.

e) Session 5 was aimed at applying the insights obtained thus far to future study situations. The imagery component focused on a future (ideally, near-future) learning-related event or situation that calls for behavioral change. The rescripting process for this imagined situation involved identifying the expected blocks to such behavior, a dialogue between the ego-states involved (e.g., the perfectionistic or procrastinating voice vs. the healthy student voice), and finally, rehearsal of the actual behavior in imagery. The imagery work was followed by a review of the client's modes of learning and test-taking skills, with the aim of identifying behaviors that could be improved. The post-session homework involved practice in monitoring test-taking situations, again identifying alternative behaviors and cognitions.

f) Session 6 was aimed at applying the insights obtained thus far to future test-taking situations. The imagery component focused on a future (ideally, near-future) test-related situation that calls for behavioral change, but in all other respects, was similar to that used in Session 5. The imagery work was followed by a review of the entire treatment protocol's materials, with particular focus on both emotional and behavioral insights from the imagery exercises as well as the homework assignments. The post-session homework involved encouragement to continue practicing those behavior changes that proved to be most promising during the intervention.**Study design** Fig. 2 presents the study design. We used a non-concurrent multiple baseline design, with random assignment to varying baseline lengths. The study was designed with the clients' semester schedules in mind. Before the (varying) baseline periods, all participants completed two sets of measures: at recruitment (Time 0) and one week later (Time 1). Group 1 started treatment after completing weekly baseline questionnaires for two weeks (Time 2a-b), Group 2 started after completing four weeks (Time 2a-d), and Group 3 started after completing six weeks (Time 2a-f). A final baseline assessment (Time 3) was completed before the first session. During the treatment, clients completed symptom measures (namely, the STA) prior to each session; they also completed session evaluation measures (namely, the EE-SR, SAI, and SES) following each session (Time 4a–f). An immediate follow-up (Time 5) was completed following the final (6th) session. A 2nd follow-up were completed one week after the final session (Time 6). An additional delayed follow-up (Time 7) was completed six weeks (Group 1), 4 weeks (Group 2), or 2 weeks (Group 3) after treatment completion. According to the semester schedule, Time 7 was assessed in the exam period. At Time 7, most of the clients had sat for at least one exam, and the majority were still preparing for more upcoming exams. The client received recruitment, baseline and follow-up assessments online. During the treatment the client as well as the therapist answered the measures using a PC. After each session the therapist received a feedback of the clients' answers.

2.4. Measures

All measures were German or Hebrew translations of these English originals, and were back-translated to ensure consistency.

Test anxiety measures. At Times 0,1,3,5, and 7, clients completed Spielberger's Test Anxiety Inventory (TAI; 1980), a 20-item self-report measures assessing level of test anxiety using a 1 (almost never) to 4 (almost always) Likert scale. The TAI is the most frequently used measures for test anxiety (Chapell et al., 2005) and has been found to have good test–retest reliability ($r = 0.80$; Spielberger, 1980). In the current study we used its global score (a sum of all items), which showed good internal reliabilities at Time 0 (Cronbach's α (N = 59): 0.91) and in the four subsequent assessments during the study (Cronbach's α ($N = 31$): 0.80 (Time 1), 0.77 (Time 3), 0.85 (Time 5 and Cronbach's α ($N = 30$): 0.95 (Time 7)). In addition, clients' test anxiety was monitored during the BL (Times 2a-f) and treatment (Times 4a-f), and followup (Time 6 and Time 7) periods using the State Test Anxiety measure (STA; Lawrence and Williams, 2013), a brief measure consisting of two items (*I feel*

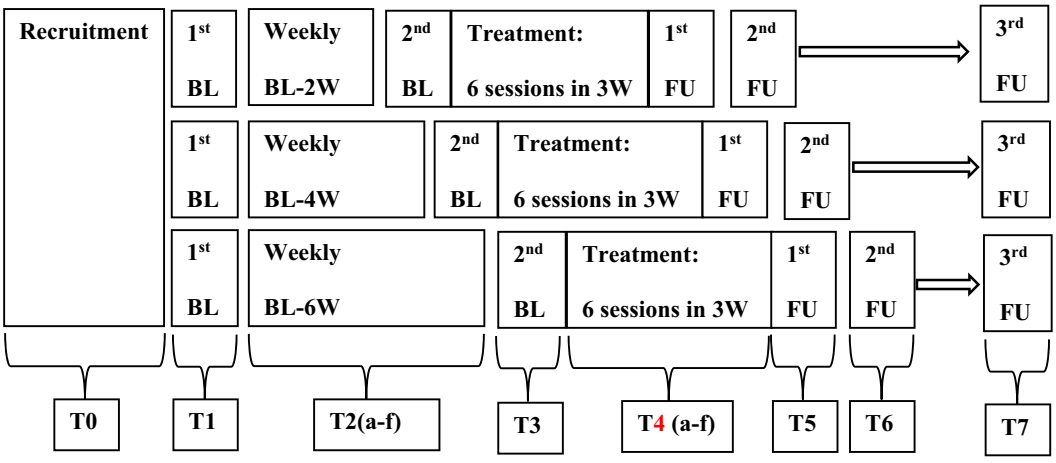

Fig. 2. Study Design. Note: BL = Baseline; FU = Follow up.

anxious about taking the upcoming tests; I feel distressed and uneasy about taking the upcoming tests) rated on 0 (completely disagree) to 6 (completely agree) Likert scale. The between- and within-person reliabilities for the scale was computed using procedures outlined by Shrout and Lane (2012); they were 0.72 and 0.84.

Emotional process measures. The Emotional Experience Self-Report (EE-SR) was administered following each session. This single item measure, from Fisher et al. (2016) asks clients to note how intense the experience of their emotions was during the session using a graphic scale ranging from 0 (*In today's session I was disconnected from my emotions*) to 100 (*In today's session I experienced my emotions vividly and fully*). Additionally, we assessed the potency of the imagery component of each session using 5 items adapted from Salter et al. (2015) observational coding system. Specifically, using a 1 to 4 scale, clients reported (a) their motivation to be involved in the imagery work; (b) their ability to stay with the image; (c) the vividness of the imagery; (d) the intensity of their emotions during the imagery work; and (e) their surprise during it. These items were averaged to create a general index tapping imagery potency. The between- and within-person reliabilities for the scale was computed using procedures outlined by Shrout and Lane (2012): they were 0.63 and 0.59, respectively.

Session and treatment evaluation. Following each session, clients completed the Session Alliance Inventory (SAI; Falkenström et al., 2015), Session Evaluation Scale (SES; Hill and Kellems, 2002; Lent et al., 2006). The SAI is a reliable yet brief measure for assessing working alliance repeatedly. It includes 6 items, each rated on a 0 to 6 Likert scale. In the current study we used composite scores created by averaging the 6 items. The between- and within-person reliabilities for the scale were computed using procedures outlined by Shrout and Lane (2012); these were 0.91 and 0.79, respectively. The SES is a 5-item measure which asks clients to report their perception of the quality of the session on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), using statements such as "I am glad I attended this session." The between- and within-person reliabilities for the scale were 0.78 and 0.70, respectively.

An overall evaluation of the quality of the intervention was assessed at Time 6 using 4 items adapted from the Credibility/Expectancy Questionnaire (CEQ; Devilly and Borkovec, 2000). Clients were asked to report on a 10-point Likert scale ranging from 1 (not at all) to 9 (very) their perception of how the therapy will help to reduce their test anxiety. The items used were: "At this point, how logical does the therapy that was offered to you seem?", "At this point, how successfully do you think this treatment will be in reducing your test anxiety?", "How confident would you be in recommending this treatment to a friend who experiences similar problems?" and "By the end of the therapy period, how much improvement in your test anxiety symptoms do you think has occurred?". The internal reliability of this scale was Cronbach's α 0.81.

2.5. Analytic strategy

One client dropped out before the delayed follow-up (Time 7) assessment; we carried-forward her value from the last available data point (following the final session). We also report the results without this client in a set of footnotes[1-4]. Descriptive statistics were calculated for all demographic data. Time 6 overall treatment evaluation, along with Time 4(a–f) session evaluation and imagery potency measures were used to address hypothesis 1 (which centered on the protocol's acceptability).

We examined whether either site (i.e., Trier vs. BIU) or group (i.e., 2- vs. 4- vs. 6-week baseline periods) differences were present in the data. As none were found, we collapsed the data across all groups and both sites. Traditional repeated measures ANOVA and post-hoc paired *t*-tests were conducted to compare the TAI scores from each time point (Time 0, 1, 3, 5 and 7) to address hypothesis 2, which centered on the protocol's efficacy. In addition, we used the reliable change index (RCI; Jacobson, and Truax, 1991) computed for the TAI. The standard deviation for collage undergraduate students (women: 13.7, men: 12.43) and the two-week test-retest reliability estimates (r_{tt} = 0.80) for this measure were obtained from its original manual (Spielberger, 1980). Finally, we also used growth models using STA responses from Times 2(a-f), 4(a-f), 6, and 7. These were estimated using a piecewise model for intercept/slope discontinuity (Hoffman, 2014):

$$STA_{tc} = \beta_{0c} + \beta_{1c}*BL - Time_{tc} + \beta_{2c}*Tx_{tc} + \beta_{3c}*Tx - Time_{tc} + \beta_{4c}*FU - 1_{tc} + \beta_{5c}*FU - 2_{tc} + r_{tc}$$

where STA_{tc} (i.e., the state test anxiety score at each assessment point t for client c) is modeled as a function of the intercept, a slope coefficient (β_{1c}) for the successive baseline assessments (Times 2a-f), a "jump" coefficient (β_{2c}) estimating the change between the end of the baseline period (Time 2f) and the beginning of the treatment (Time 4a), another slope coefficient (β_{3c}) for the successive pre-session assessments (Times 4a-f), and two more "jump" coefficients estimating the change between the end of treatment and the second followup (β_{4c}) and the one between the second and third followup (β_{5c}).

To test hypothesis 3 (which addressed the associations between process variables) we used hierarchical linear models

In the first of these models, SAI_{tc} (i.e., the session alliance inventory score on session t for client c) was modeled as a function of both the client's average rating of potency for imagery, as well as the rating for potency of the current session's imagery work:

$$SAI_{tc} = \gamma_{00} + \gamma_{01}*Avg.\ Imagery_{c} + \gamma_{10}*Imagery_{tc} + u_{0c} + r_{tc}$$

In the second and third models, EE-SR and SES served as the outcomes; in all other respects, these models were identical to the second model.

3. Results

3.1. Hypothesis 1: acceptability of the protocol

At the second follow up (Time 6), clients reported that they had found the treatment to be feasible using the CEQ, a measure which inquires about how logical, successful, and recommended the treatment is. Overall, clients found the therapy to be acceptable, with an average CEQ rating of 5.8 (range 3.3–8.8; SD = 1.4, on a scale anchored by 1 and 9). Similarly, mean scores from the Session Evaluation Scale (SES: session one to session six) showed reported quality ratings to be high (M = 4.31, SD = 0.6). The average ratings increased over the course of the protocol's sessions (session 1: 4.2, sessions 2: 4.3, session 3: 4.2, Session 4: 4.3, session 5: 4.4, and sessions 6: 4.5). A post-hoc paired sample *t*-test showed that this was a significant increase from session 1 to session 5 ($t(30) = -2.3$, $p < .05$, $d = -0.40$) and session 6 ($t(30) = -3.4$, $p < .01$, $d = -0.56$).

The clients' ratings of the potency of the imagery component using the ImRs coding framework from session 1 to session six were also high (M = 3.0, SD = 0.5). The average ratings fluctuated somewhat from session to session, with the lowest rating (session 1: 2.8) in the session in which imagery was focused solely on safe-place images, and with higher ratings (session 2: 2.9, session 3: 3.2, session 4: 3.1, session 5: 2.9 and session 6: 3.1) for sessions with exploratory or rescripting imagery. A post-hoc paired sample *t*-test showed that this was a significant increase from session 1 to session 3 ($t(29) = -4.0$, $p < .01$, $d = -0.81$).

3.2. Hypothesis 2: treatment outcome

A one-way repeated measure ANOVA with time as the independent variable and clients' reports of TA as the dependent variable showed

that the treatment reduced clients' test anxiety significantly ($F(4,120) = 10.16$, $p < .001$, $\eta^2 = 0.25$).[1] Post-hoc paired sample *t*-tests showed that TAI scores dropped significantly from recruitment to delayed follow-up (Time 0 to Time 7: $t(30) = 4.0$, $p < .001$, $d = -0.75$),[2] baseline to delayed follow-up (Time 1 to Time 7: $t(30) = 4.4$, $p < .001$, $d = -0.84$)[3] and pre-intervention to delayed follow-up (Time 3 to Time 7: $t(30) = 3.8$, $p < .001$, $d = -0.72$)[4]. Fig. 3 shows the mean TAI scores over time with confidence intervals and Table 2 presents the complete results of the post-hoc paired sample *t*-tests. A comparison of average pre-treatment scores (from Time 0, Time 1, Time 3) to post-treatment scores (from Time 5 or Time 7) using a post-hoc *t*-test showed that TAI scores dropped significantly ($t(30) = 3.9$, $p < .001$, $d = -0.59$).

Table 3 presents the RCI for the TAI from pre-treatment (Times 0, 1, and 3) to delayed follow-up (Time 7). One participant declined from Time 0 (recruitment) to delayed follow-up; the majority remained unchanged, but approximately 25% experienced positive reliable change by Time 7.

The discontinuity growth model for STA scores (see Fig. 4) revealed that during the baseline period, clients did not undergo any significant change (Times 2a-f;b = −0.001, SE = 0.01, $p = .89$); they manifested a significant rise in symptoms at the beginning of treatment (between Time 2f and Time 4a;b = 0.36, SE = 0.17b = , $p < .05$) as well as a significant decline during treatment (Times 4a-f;b = −0.14, SE = 0.04, $p < .001$) and a further decline between the second and third follow ups (Times 6 and 7;b = −0.54, SE = 0.17, $p < .01$).

3.3. Hypothesis 3: treatment process

Hierarchical linear models were estimated to examine the associations between the hypothesized process variable (potency of imagery work) and three session-level outcome variables (SAI, EE-SR, and SES). With session alliance (SAI) as the outcome, both the client average and the same session ratings of imagery potency were significant predictors (b = 0.78, SE = 0.29, $p < .05$; b = 0.37, SE = 0.08, $p < .001$, respectively). The same was true for the reported emotional experience (EE-SR; b = 19.18, SE = 9.0, $p < .05$; b = 14.87, SE = 3.05, $p < .001$) and the overall session evaluation (SES; b = 0.90, SE = 0.15, $p < .001$; $b = 0.42$, SE = 0.07, $p < .001$).

4. Discussion

This pilot study examined the acceptability and effectiveness of a newly developed six-session imagery-based protocol for the treatment of test anxiety. It also explored the processes that contribute to these. It joins the recently published study by Reiss et al. (2017) demonstrating similar effects for imagery rescripting in group therapy for TA. Clients found the treatment to have a logical structure, and rated the sessions (in general) as effective and the imagery component of the sessions (in particular) as potent. Test anxiety levels were reduced significantly from pre- to post-treatment (Times 0,1,3 to Time 7), as evidenced by decrements in the TAI. This improvement in the TAI was reliable for nearly one quarter of the clients from pre-treatment to delayed follow-up, with no client manifesting a reliable deterioration from baseline to delayed follow-up. Whereas TA measured with the STA showed a significant decline between Time 6 and Time 7, no significant change occurred between the end of the treatment (Time 4f) and the second follow-up (Time 6), in contrast to what we had expected. Measurement-related issues (e.g., the reliance on the STA index which includes only two items) may be at play, as may be the relatively brief time for practice (one week). Finally, process analyses revealed that the putative active mechanism of the imagery work (namely, the potent activation of affect during any session) was indeed predictive of same- session evaluations, experienced emotion, and therapeutic alliance ratings, which were all higher. Using HLM allowed us to ascertain that these associations cannot be explained by between-subject differences but rather that they are specific to within-session processes, as we hypothesized. While the patterns of findings reported here are novel, it is clear that results must be interpreted with caution and require replication.

These results provide promising evidence for the acceptability, feasibility, and utility of using imagery-based methods in the treatment of test anxiety. Nonetheless, this study is not free of limitations. For one, it involved a relatively small sample size, leading to weaker power to detect treatment effects. A partial remedy for this limitation comes from the use of repeated measures both at baseline and during the treatment, yet a replication of this study with a larger sample is certainly needed to gain greater confidence in the results. Another limitation was that some of the clients did not have any exam scheduled after the treatment. Some clients completed their exams before the delayed follow-up assessments. We wanted to have a uniform time period elapse between the treatment and these follow-ups (1 week until the first delayed follow-up, and 2, 4, or 6 weeks [depending on group assignment] until the second delayed follow-up). Though we were able to request yet another (post-exam) report from some of the clients (those from the Trier site, which can be thought of as Time 8), the variability in the time lags and the incompleteness of such data rendered it less useful. Importantly, these data suggested some further reduction in symptoms from Time 7 to Time 8, though not a significant one ($t(22) = 1.73$, $p = .098$, $d = -0.39$). Therefore, we conclude that the further decrease in TA between Time 5 and Time 7 is due mainly to the treatment (e.g. more time to practice) rather than the passage of the exams. Future studies could avert this problem by enrolling only participants who share a more fixed academic schedule.

In addition to the timing of the treatment vis-à-vis the students' exams (and thus, post-exam assessments), we have come to realize that future implementation of this protocol may benefit from re-considering the intensity of sessions. In the current study, we conducted the six sessions over the span of three weeks, mostly for scheduling reasons. The feedback from the therapists and supervisors was that this felt somewhat rushed, often leaving little time for the clients to practice skills or complete tasks discussed in the sessions. It is also possible that the non-significant findings between the recruitment (Time 0), baseline (Time 1) and pre-treatment (Time 3) and the first follow-up (Time 5) assessments are due to the high intensity of the protocol. Moreover, the Time 5 assessment, which was obtained in the room right after the completion of the last session, may have not allowed the clients enough time to implement any new skills or insights learned in that session.

Importantly, the overall effect size (Time 0 to Time 7: $d = 0.84$) we reported is within the range of the more effective approaches included in Ergene's (2003) meta-analysis (0.80–1.22), but does not exceed them. This may raise the question whether adding imagery work improves on existing treatments addressing TA. Notably, Ergene pointed out that their results should be interpreted with caution, as most of the groups' effects failed to be homogeneous and a significant amount of variance remained unexplained. However, to fully answer this question, future work will require conducting a randomized control study.

Finally, the clients recruited into the study were not assessed using structured clinical interviews, and thus their diagnostic profiles are not fully known to us. Future work should seek to obtain additional diagnostic information which could inform us about the extent to which these outcome results are generalizable to a broader patient population.

Despite these limitations, the findings of this study add to the growing body of knowledge on possible interventions for test anxiety. Several interventions utilizing various treatment strategies are used for TA (Ergene, 2003). The fact that combined approaches shows the

[1] $F(4,116) = 9.8$, $p < .001$, $\eta^2 = 0.25$.
[2] $t(29) = 3.9$, $p < .01$, $d = 0.86$.
[3] $t(29) = 4.2$, $p < .001$, $d = 0.73$.
[4] $t(29) = 3.8$, $p < .01$, $d = 0.75$.

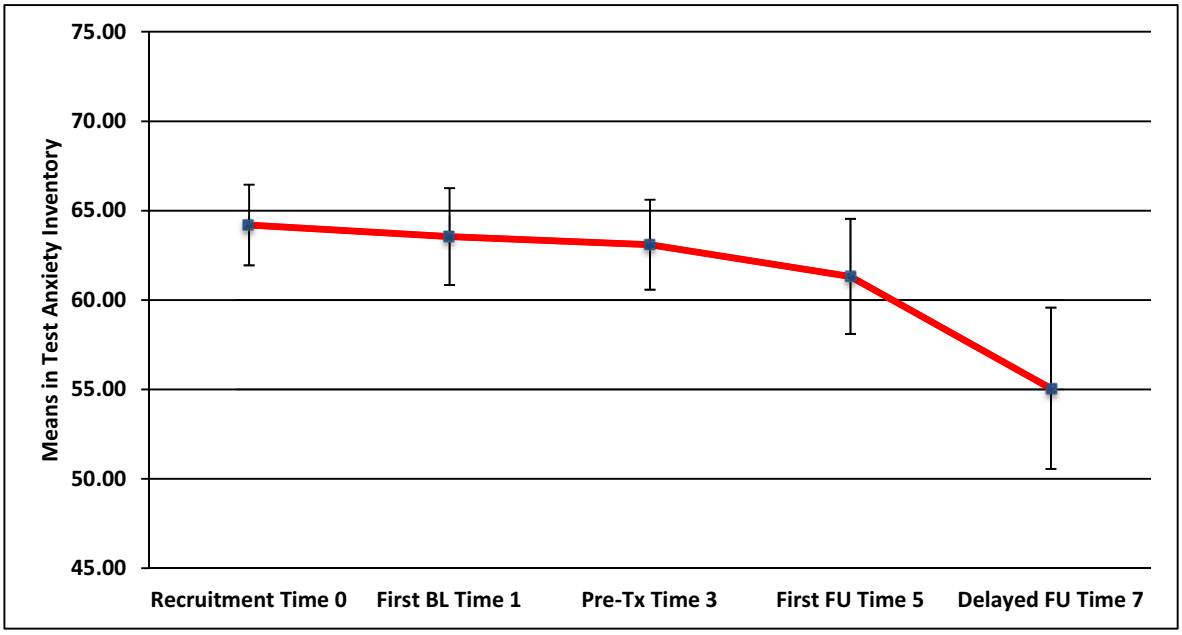

Fig. 3. Mean Test Anxiety Inventory (TAI) Score over Time with confidence intervals ($N = 31$).

Table 2
Complete results of the post-hoc paired sample *t*-tests for the Test Anxiety Inventory (TAI) on 5 assessments.

Pair	Mean	SD	*t* Statistic	*df*	*p* Value	*Cohens d*
T0 - T1	2.4	9.3	1.5	35	.138	−0.260
T0 - T3	1.1	6.9	0.9	30	.384	−0.159
T0 – T5	2.9	8.2	1.9	30	.061	−0.434
T0 – T7	9.1	12.7	4.0	30	<0.001	−0.746
T1 - T3	0.45	5.6	0.45	30	.655	−0.081
T1 – T5	2.2	7.8	1.58	30	.124	−0.315
T1 – T7	8.5	10.8	4.4	30	<0.001	−0.836
T3 – T5	1.8	8.25	1.2	30	.241	−0.218
T3 – T7	8.03	11.8	3.8	30	.001	−0.716
T5 – T7	6.26	11.0	3.17	30	.004	−0.587

Table 3
Reliable change index (RCI) analyses with Test Anxiety Inventory (TAI) scores.

	Time 0 to time 7 $d = 0.87$	Time 1 to Time 7 $d = 0.76$	Time 3 to Time 7 $d = 0.74$
TAI (RCI for women = 16.98; RCI for men = 15.41)			
Total N	31	31	31
Improvement	8 (25.8%)	6 (19.4%)	6 (19.4%)
No change	22 (71.0%)	25 (80.6%)	25 (80.6%)
Deterioration	1 (3.2%)	0 (0%)	0 (0%)

highest effects leads to the suggestions that TA is more than a deficit of study and test taking skills as wells as a deficit of knowledge (Ergene, 2003). There is a lack of studies, which investigate in the putative underlying mechanisms behind effective TA programs. Social phobia studies, which investigate in the effectiveness of ImRs in combination with cognitive restructuring discuss the change of the meaning of the early traumatic experience as core mechanisms (Lee and Kwon, 2013). This goes in line with the results of Reiss et al. (2017) as well as with the results of this study. This study broadens the range of target symptoms and disorders for which emotion-focused imagery-based interventions appear to have utility. We are excited to continue and hone this intervention, and while doing so, to deepen our understanding about the possible mechanisms at work here. For example, given the consistent evidence for a significant role for the therapeutic bond in psychotherapy (Rubel et al., 2017; Lutz et al., 2015), we hope to explore the possibility that the imagery work exerts some of its effects on the alliance and on outcome- In this study, we linked the potent imagery work to session-level alliance. So far, we did not test for mediation. It is possible that the effect of alliance is mediated through a concomitant emotional activation in *the therapist*. If that is the case, such activation should be observable in the *therapist's* physiological reactions, as well as in synchrony in such reactions between clients and therapists. Such a finding would be consistent with the evidence that client-therapist concordance (i.e., synchrony) is tied to clients' perception of their therapists' empathy (e.g., Marci et al., 2007)—and through it, to positive therapy outcomes.

Contributors

Jessica Prinz was responsible for the concept, data collection, data analyses, interpretation, and writing. Eran Bar-Kalifa contributed to the concept, coordinated the data collection as well as analyses and writing. Eshkol Rafaeli contributed to the concept as well as writing and editing. Haran Sened contributed to the concept as well as writing. Wolfgang Lutz contributed to the concept as well as writing and editing.

Funding

The authors received no funding from an external source. The authors declare no conflict of interest.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

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

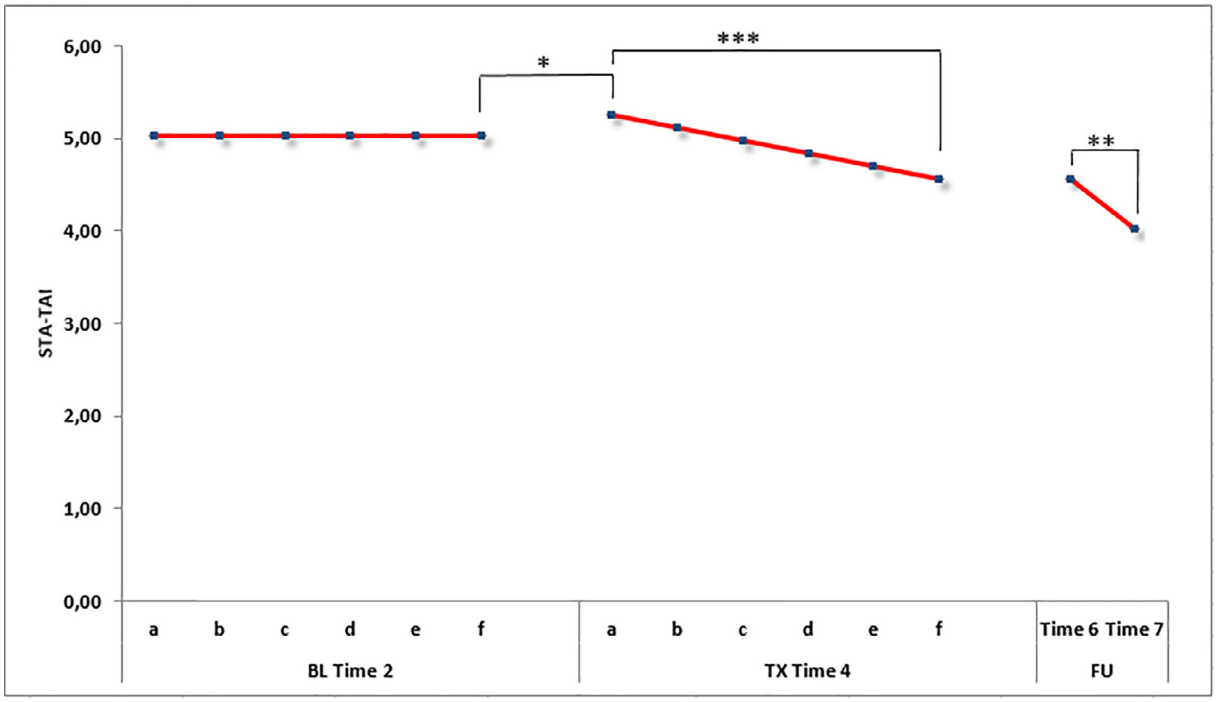

Fig. 4. Growth model of state test anxiety (STA).

Acknowledgement

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2018.10.091.

References

Arntz, A., Weertman, A., 1999. Treatment of childhood memories: Theory and practice. *Behav. Res. Ther.* 37 (8), 715–740.
Arntz, A., 2014. Imagery rescripting for personality disorders: healing early maladaptive schemas. In: Thoma, N.C., MacKay, D. (Eds.), Working with Emotion in Cognitive-behavioral Therapy: Techniques for Clinical Practice. Guilford Publications.
Brown, L.A., Forman, E.M., Herbert, J.D., Hoffman, K.L., Yuen, E.K., Goetter, E.M., 2011. A randomized controlled trial of acceptance-based behavior therapy and cognitive therapy for test anxiety: a pilot study. *Behav. Modif.* 35 (1), 31–53.
Chapell, M.S., Blanding, Z.B., Silverstein, M.E., Takahashi, M., Newman, B., Gubi, A., McCann, N., 2005. Test anxiety and academic performance in undergraduate and graduate students. *J. Educ. Psychol.* 97 (2), 268–274.
Devilly, G.J., Borkovec, T.D., 2000. Psychometric properties of the credibility/expectancy questionnaire. *J. Behav. Therapy Exp. Psychiatry* 31 (2), 73–86.
Eum, K., Rice, K.R., 2011. Test anxiety, perfectionism, goal orientation, and academic performance. *Anxiety Stress Coping* 24 (2), 167–178.
Ergene, T., 2003. Effective interventions on test anxiety reduction a meta-analysis. *School Psychol. Int.* 24 (3), 313–328.
Falkenström, F., Hatcher, R.L., Skjulsvik, T., Larsson, M.H., Holmqvist, R., 2015. Development and validation of a 6-item working alliance questionnaire for repeated administrations during psychotherapy. *Psychol. Assess.* 27 (1), 169.
Fehm, L., Fydrich, T., 2011. Prüfungsangst. [Test Anxiety]. Hogrefe Verlag.
Fisher, H., Atzil-Slonim, D., Bar-Kalifa, E., Rafaeli, E., Peri, T., 2016. Emotional experience and alliance contribute to therapeutic change in psychodynamic therapy. Psychotherapy 53 (1), 105–116.
Hill, K.T., Wigfield, A., 1984. Test anxiety: a major educational problem and what can be done about it. *Element. School J.* 85 (1), 105–126.
Hill, C.E., Kellems, I.S., 2002. Development and use of the helping skills measure to assess client perceptions of the effects of training and of helping skills in sessions. *J. Counsel. Psychol.* 49 (2), 264–272.
Hoffman, L., 2014. Longitudinal Analysis: Modeling within-Person Fluctuation and Change. Routledge.
Holmes, E.A., Arntz, A., Smucker, M.R., 2007. Imagery rescripting in cognitive behaviour therapy: images, treatment techniques and outcomes. *J. Behav. Ther. Exp. Psychiatry* 38, 297–305.
Holmes, E.A., Mathews, A., 2010. Mental imagery in emotion and emotional disorders. *Clin. Psychol. Rev.* 30 (3), 349–362.
Holmes, E.A., Mathews, A., 2005. Mental imagery and emotion: a special relationship? Emotion 5 (4), 489–497.
Holm-Hadulla, R.M., Hofmann, F.H., Sperth, M., Funke, J., 2009. Psychische Beschwerden und Störungen von Studierenden [Psychological complaints and mental disorders of students]. Psychotherapeut 54 (5), 346–356.
Ji, J.L., Heyes, S.B., MacLeod, C., Holmes, E.A., 2016. Emotional mental imagery as simulation of reality: fear and beyond – a tribute to Peter Lang. *Behav. Ther.* 47 (5), 702–719.
Jacobson, N.S., Truax, P., 1991. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J. Consult. Clin. Psychol.* 59 (1), 12–19.
Lawrence, J.S., Williams, A., 2013. Anxiety explains why people with domain-contingent self-worth underperform on ability-diagnostic tests. *J. Res. Pers.* 47 (3), 227–232.
Lee, S.W., Kwon, J.H., 2013. The efficacy of imagery rescripting (IR) for social phobia: a randomized controlled trial. *J. Behav. Ther. Exp. Psychiatry* 44, 351–360.
Lent, R.W., Hoffman, M.A., Hill, C.E., Treistman, D., Mount, M., Singley, D., 2006. Client-specific counselor self-efficacy in novice counselors: relation to perceptions of session quality. *J. Counsel. Psychol.* 53 (4), 453.
Lutz, W., Rubel, J., Schiefele, A.K., Zimmermann, D., Böhnke, J.R., Wittmann, W.W., 2015. Feedback and therapist effects in the context of treatment outcome and treatment length. *Psychother. Res.* 25 (6), 647–660.
Marci, C.D., Ham, J., Moran, E., Orr, S.P., 2007. Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *J. Nerv. Mental Dis.* 195 (2), 103–111.
Morris, L.W., Liebert, R.M., 1970. Relationship of cognitive and emotional components of test anxiety to physiological arousal and academic performance. *J. Consult. Clin. Psychol.* 35 (3), 332.
Nilsson, J.-E., Lundh, L.-G., Viborg, G., 2012. Imagery rescripting of early memories in social anxiety disorder: an experimental study. *Behav. Res. Ther.* 50 (6), 387–392.
Rafaeli, E., Maurer, O., Lazarus, G., Thoma, N.C., 2016. The self in schema therapy. In: Kyrios, M., Bhar, S., Doron, G., Mikulincer, M., Moulding, R., Nedeljkovic, M. (Eds.), The Self in Understanding and Treating Psychological Disorders. Cambridge University Press, Cambridge, UK.
Rafaeli, E., Maurer, O., Thoma, N.C., 2014. Working with schema modes. In: Thoma, N.C., McKay, D.M. (Eds.), Working With Emotion in Cognitive-Behavioral Therapy: Techniques for Clinical Practice. Guilford, New York.
Reiss, N., Warnecke, I., Tolgou, T., Krampen, D., Luka-Krausgrill, Rohrmann,S., 2017. Effects of cognitive behavioral therapy with relaxation vs. Imagery rescripting on test anxiety: a randomized controlled trial. *J. Affect. Disord.* 208, 483–489.
Rubel, J., Rosenbaum, D., Lutz, W., 2017. Patients' in-session experiences and symptom change: session-to-session effects on a within- and between-patient level. *Behav. Res. Ther.* 90, 58–66.
Safran, S., Otto, M., Sprich, S., Winett, C.L., Wilens, T., Biederman, J., 2005. Cognitive-behavioural therapy for ADHD in medication-treated adults with continued symptoms. *Behav. Res. Ther.* 43, 831–842.
Salter, C., Parker, E., Brown, G., El Leithy, S., Wheatly, J., 2015. Imagery Rescripting Coding Framwork and Manual. DOI: 10.13140/RG.2.1.4420.5282.
Shrout, P., Lane, S.P., 2012. Psychometrics. Handbook of Research Methods for Studying Daily Life. Guilford Press.

Spielberger, C.D., 1980. Test Anxiety Inventory: Sampler Set: Manual, Test Scoring:" Test Attitude Inventory": Preliminary Professional Manual. Mind Garden.
Wild, J., Clark, D.M., 2014. Experimental Exercises and Imagery Rescripting in Social Anxiety Disorder: New Perspectives on Changing Beliefs. In: Thoma, N.C., McKay, D.M. (Eds.), Working with Emotion in Cognitive-Behavioral Therapy: Techniques for Clinical Practice. Guilford: New York.
Zeidner, M., 1998. Test Anxiety: The state of the Art. Springer Science & Business Media.
Zeidner, M., Matthews, G., 2010. Anxiety 101. Springer Publishing Company.
Zeidner, M., 2014. Test anxiety. The Wiley Handbook of Anxiet Disorders. pp. 581–595.